Universidad Autónoma de Madrid

Departamento de Biología Molecular y Celular

Role of CTCF in heart development: 3D genomic structure and regulation of the IrxA locus

Melisa Gómez Velázquez

Madrid, 2016

Departamento de Biología Molecular y Celular

Facultad de Ciencias

Universidad Autónoma de Madrid

Role of CTCF in heart development: 3D genomic structure and regulation of the IrxA locus

Melisa Gómez Velázquez

Químico Bacteriólogo Parasitólogo

Doctoral thesis directed by Dr. Miguel Manzanares Fourcade

CNIC, Madrid, 2016

This work was performed in Miguel Manzanares´ Lab in the Cell and Developmental Biology Area at the Centro Nacional de Investigaciones Cardiovasculares (CNIC) in Madrid.

A mi familia y amigos

AKNOWLEDGMENTS

Agradezco a Miguel Manzanares por dejarme realizar este trabajo y ser parte en su laboratorio, donde me he encontrado con mucha gente maravillosa. Por mandarme a tantos cursos y congresos, dejar que me desapareciera temporadas largas para ir a mi casa y en general por todo el apoyo y la paciencia recibidos en estos 7 años. Por tener siempre la puerta abierta para resolver dudas o discutir cualquier cosa. Por nunca darse por vencido con este proyecto y buscar siempre la manera de tirar para adelante. Por estar presente todo este tiempo. Aprendí muchas cosas que estoy segura que me serán útiles más adelante. Disfruté mucho trabajar al lado de una persona a la que le gusta pensar y que tiene una curiosidad infinita por la ciencia.

Este trabajo no hubiera sido posible sin la ayuda de muchas personas. Así que por eso, quiero agradecer a todas las personas directamente implicadas en este proyecto de tesis. Antes que a nadie, a Claudio Badía porque sin él me habría llevado el doble de tiempo realizar este trabajo o probablemente no lo habría acabado. Gracias por estar siempre listo para trabajar, por tu excelente organización y dejar que me despreocupe de las colonias. Aprecio un montón tu energía, esfuerzo, interés, dedicación y cariño que has puesto en este proyecto. Gracias por aguantarme estos años, por los consejos y las palabras de aliento, sobre todo al final de la tesis. Por las aventuras vividas fuera del laboratorio y por tu amistad. No tengo equipo de futbol, pero que sepas que apoyo al Atlético de Madrid. A Isa que es indudablemente un ejemplo a seguir dentro y fuera del laboratorio, por sus consejos y constantes ánimos que agradecí más de una vez. A Eva Fernandez por su ayuda en el proyecto, por su paciencia para responder mis preguntas y por su forma de ser tan llena de energía y entusiasmo que nos contagiaba a todos. Ah, y por introducirnos a la Feria de Sevilla. Alberto de Genómica por responder todas y cada una de las dudas (que no fueron pocas) y por ser un gran apoyo en el mundo de la secuenciación. A Carlos Torroja por resolver mis dudas y ayudarme con R. A Frank y a Merche por cuidar de las colonias de ratón y estar al pendiente de nosotros todo el tiempo. Es agradable contar con su apoyo que ayuda a que el trabajo vaya sobre ruedas. Gracias a Eva Alonso por introducirme al mundo de la microinyección y de las transferencias. A Juan Tena por el análisis de los 4C y por resolver dudas que iban surgiendo sobre la marcha. A Jose Luis Gómez Skarmeta por ser un entusiasmado del proyecto y de la ciencia en general. Gracias a todos por su contribución, compromiso y entusiasmo al trabajar. De todos aprendí algo en menor o mayor medida.

A todos los miembros del laboratorio de Miguel Manzanares: antiguos, pasajeros y actuales. Gracias a Teresa Rayón por recibirme e integrarme (junto con Susana) en el laboratorio y en el departamento, por sus constantes ánimos cuando las cosas parecían no funcionar, en especial los días de microinyección o más bien los días de

abrir ratonas con úteros vacíos o teñir embriones sin éxito alguno. Por escuchar, por ser como las proteínas de choque térmico, por su pasión por la ciencia y por su increíble entusiasmo para hacer "todinchi". A Susana Cañón por hacerme sentir bienvenida al grupo cuando llegué, por enseñarme a cultivar embriones de pollo y guiarme en el proyecto de fin de master. Por sus consejos a lo largo de los años, por su sentido del humor y por tener iniciativa infinita a la hora de sugerir planes. Tu compañía siempre es agradable. A Julio por ser como es, después de conocerte y acostumbrarme a tus peculiares opiniones y comentarios, uno se da cuenta que hacen falta más personas como tú en la vida. Por tu iniciativa, entusiasmo, por ser un animado de la vida y por tu eterna curiosidad acerca de todo en general. Gracias también por estar ahí para dar ánimos cuando lo necesité. A Sergio Menchero, mi compañero de ordenador, por su humor (a veces de abuelo gruñón, sobre todo cuando me ayudó con las qPCRs) e infinita paciencia tanto dentro como fuera del labo. Tu forma de ser, no sé bien como, ayuda a que cualquier grupo fluya mejor. Gracias por escuchar todas y cada una de las veces que quería decir algo, así fuera irrelevante. Por preocuparse por mí, por enseñarme a bailar Sevillanas, y sobre todo por aguantarme, junto con Jesús, al final de la tesis. A Jesús por aguantarme, darme ánimos, por presentarnos a Jose, por ser un cadillo, recordarme que todos llevamos un niño dentro, y por introducir su energía positiva en el grupo. A Mariajo por su intensidad, por alimentarme y por la alegría con que te saluda por las mañanas. A Aurora por ser tan positiva y guiarnos por Berlín. Tuve dos estudiantes: Alba y Agustín, gracias por dejarme compartir lo que sabía durante los veranos. A Raquel, Gonzalo y Elena por dejarme compartir el mundo del 4C. A Alba, la más reciente adquisición del grupo por explicarme bioinformática, por su iniciativa y por trabajar con entusiasmo en el proyecto. Al resto del grupo Manzanares: Beatriz, Cristina, Luis, Inma y Maria. Gracias a todos por hacer el día a día mucho más agradable y llevadero y por su disposición a ayudar a los demás.

A Simon Bartlett por ayudarme y orientarme para escribir en inglés, a Kenneth McCreat por re-escribir esta tesis, creo que no soy la única que te lo agradece ☺. A la unidad de transgénesis, en especial a David por los consejos para microinyectar. También a los miembros de todo el departamento de... ha cambiado de nombre, integrantes y ubicación en estos 7 años. Toda esta gente, junto con los integrantes nuevos de los grupos y el resto del centro, han enriquecido la visión de la ciencia al tocar temas tan diversos. Agradezco haber estado tantos años aprendiendo de temas tan lejanos de lo mío. También disfruté mucho de los retreats que se organizaron. Gracias también por darme apoyo, consejos, prestarme reactivos y pasarme protocolos. Fue increíble llegar a un lugar donde la gente estaba siempre dispuesta a ayudarte: Alberto, Jesús Chamorro, Gaetano, Guille, Álvaro, Luis, Belén, Bea, Pablo

Gómez del Arco (se me olvida gente, seguro, porque el departamento era grande y han pasado muchas personas a lo largo de estos años), por ser fuente de conocimiento y resolverme dudas, a los Pomperos en general por impregnar alegría en el departamento. Pero sobre todo por las cañas y las múltiples actividades extra escolares que se organizaron a lo largo de estos años, a través de las cuales fui conociendo Madrid y sus costumbres.

Quiero mencionar que la infraestructura del centro es algo que ha facilitado el trabajo diario. Ese apoyo que está ahí y que solo te haces consciente cuando se te rompe un aparato o se te traba el ordenador (gracias Alicia), por ejemplo. La gente encargada de logística y compras del departamento Teresa Casaseca, Mary, Marta Ramón y en su día Sandra Cillero, así como los lab managers Ángel y Beatriz. También a los de recursos humanos, René, Carmen y Sara, que me ayudaron siempre para mis trámites de extranjería. Las chicas de la biblioteca Irene y Alicia. Gracias, porque al realizar su trabajo, muchas veces facilitaron el mío.

A Sergio Casas por ser una luz cuando las cosas se veían oscuras y por su paciencia al escucharme. Por darme consejos que me son útiles hasta la fecha. Y por haber sido un apoyo siempre que lo necesité. Simplificas todos los problemas y siempre es agradable hablar contigo. Eres una persona muy resolutiva, comprensiva, paciente y optimista, ojalá cada laboratorio tuviera alguien como tú.

A mi familia en Hungría: Ixchelt, Gerardo (con todo y lo carrilla que eres) y Leo. Por recibirme en su casa y hacerme sentir como en familia, por los viajes compartidos y las risas y recuerdos de esos viajes. ¿Como olvidar Sarospatak (99% segura de que así no se escribe): “Taxi driver, no home”, y el sushi más caro de la historia, entre muchas otras anécdotas de las que me acuerdo y me todavía me río. También por ser mí consulta privada a distancia. Son un gran ejemplo de lo que es el verdadero trabajo en equipo, la paciencia y la perseverancia.

A Marina Peralta por compartir su comida (no siempre voluntariamente =S), por escucharme y aconsejarme con su sentido del humor para que mirara más allá de mis narices, por pasar su tiempo libre conmigo, aguantarme, por compartir viajes con sopas de cebolla y sobre todo por ser mi amiga. A Julie por ser excelente persona en general. Tenerte como compañera de piso aumentó mi calidad de vida fuera del labo. Te admiro a nivel laboral, personal y en todos los niveles en que se puede admirar a una persona. A Vero por su extraño sentido del humor, a Briggite por escucharme y por sus conversaciones, a Mónica Gzz por su amistad, consejos científicos y por presentarme a Liz.

Agradezco a Massi por ser de esas personas que te empujan y ayudan a cruzar tus límites, por introducirme al mundo de la salsa, por escucharme y sobre todo por tus consejos. Todos y cada uno de los que seguí, tuvieron un efecto positivo en mi vida. A Mónica López por su ayuda en general y valiosa contribución para que yo saliera adelante. A Alberto, el abogado del diablo, que siempre tiene una visión más amplia de las cosas que te ayudan a tener en cuenta otras perspectivas, y que al igual que Massi, te escucha y te empuja. A Pablo por entender mi sentido del humor y compartir risas, por escuchar, por estar al pendiente de mí, porque disfruto mucho de tu compañía y porque no recuerdo un momento aburrido a tu lado. A Ale por tu determinación admirable a la hora de hacer cualquier cosa, porque de tu iniciativa y visión del mundo aprendí a exprimir mucho más cada experiencia y situación, y por tu apoyo en los años de la tesis. A Daya por hacer de cada evento una fiesta y por tu actitud positiva ante todas las adversidades. A Liz Quintero por tu sentido del humor, consejos y por responder honestamente a mis preguntas. A Nayeli por tu amistad a distancia y por sus consejos. Sin saberlo, han contribuido a mi desarrollo personal, tengo otra perspectiva de muchas cosas, soy consciente de detalles que antes me pasaban desapercibidos. La vida en Madrid no hubiera sido igual sin ustedes.

A María Galardi por su maravilloso timing y por introducirme al mundo del peloto patinador de Impulso Urbano. El mundo no es el mismo una vez que se aprende a patinar y menos cuando te das cuenta que en realidad no sabías. A Celia y Aurelie por ser mis compañeras y amigas sobre ruedas, las rutas por Madrid son mejores cuando las hago con ustedes.

A Carlos por traer un pedacito de Mty y de la FCB a Madrid. A mis contemporáneos miembros del L8 (en especial a Ricardo) y a la Doctora Diana, por introducirme al mundo de la biología del desarrollo. A Pedro y a Diego por proveerme de experiencia y conocimiento que me resultó muy útil en los años de tesis. Y a Manolo por incitarme a hacer el doctorado.

A Briane por ser maravillosa como ella sola. You are fun as hell!. I was really glad you came into the lab and hence into my life, I enjoyed your stories, your company and just hanging out with you. You are a great person!. A Noelia, la mejor compañera de karaoke que improvisa y se pone a la altura de la situación, aun cuando no conoce la canción. Por su adicción al baile y por hacerme reír tanto con su "cuando pongan una que no me guste, nos vamos". A Ghislaine porque irradia energía positiva y la esparce a su alrededor. A Patri porque con ella me he reído muchísimo. A Cris del Carmen por sus conversaciones, consejos y apoyo en la última etapa de la tesis. A Izaskun por su amistad. A Lao y a Dimitris por sus constantes ánimos de que esto pronto acaba. A Luis

de Rui por sus conversaciones de todo un poco y por sonreír aun después de estar toooodo el día trabajando sin parar. A Joan por los ánimos en los últimos días de la escritura.

A Cris Gtzz, Héctor, Rocío y Dani por los buenos momentos en viajes, reuniones, fiestas, cumpleaños y el día a día en el laboratorio.

Al apoyo a distancia: Zeming and Roosevelt, thanks for keeping in touch after all this years, cheering up and be willing to listen to me. A Karina por ser mi amiga y escucharme desde hace años, muchos antes de la tesis y por compartir aventuras conmigo. A Arturo (a-mooor) por hacerme reír y por sus conversaciones. A Lupita por ser un oído con patas y una fuente de consejos y sabias reflexiones, por sonreírle a las adversidades, por la cantidad de energía que tienes, y por el cuidado y cariño con que haces todo. A Rodo por su amistad. A Delia porque irradia buena vibra y porque siempre que la veo, tiene una sonrisa para dar. A Deya por su apoyo años atrás. A Soraya escucharme, darme consejos, por seguir siendo mi amiga. Gracias a todos ustedes, que a pesar de la distancia y los años siguen siendo parte de mi vida.

Por último a mi familia, mis dos hermanas: Karina y Jéssica y mis papás. Me alegro de no ser hija única y de que sean parte de mi vida. Gracias a mis papás por esmerarse para que tuviéramos las mejores oportunidades y herramientas posibles. Por apoyarme a pesar de no entender exactamente qué es lo que hago, con la tesis y con muchas otras cosas a lo largo de mi vida. A mí papá por inculcarme el gusto por el deporte, por sus consejos y ánimos constantes sobre todo en situaciones difíciles. A mi mamá por preocuparse por mis cultivos de pollos (aun cuando ya trabajaba con ratones ☺), por su esfuerzo para ayudarme a resolver situaciones que no le eran para nada familiares y por estar al pendiente de mi todo el tiempo. Pero sobre todo gracias por hacerme sentir que si todo falla, no importa, porque sé que tengo un sitio a donde puedo regresar y sé que siempre van a estar ahí dispuestos a ayudarme como sea y dónde sea. Con el paso de los años, más me doy cuenta de todo lo que me han enseñado y de lo que he aprendido de ustedes. Los quiero mucho!!!

SUMMARY

The cell needs to exert a tight control over gene expression in order for the information to flow properly in time and space. To achieve this, the information stored in the genome must be highly organized for its correct deployment. There is increasing evidence that suggests that the three dimensional (3D) structure of the genome plays a major role in its organization and function. Different architectural proteins that bind chromatin are key players in 3D genome structure, and among these, CTCF has been described to have a major role in this process. In this work we aimed to explore the role of 3D genome structure mediated by CTCF in the developing heart. We find that *Ctcf* is essential for heart development, and that its loss gives rise to gross cardiac malformations that eventually lead to embryonic death. In order to gain a global view of the impact of *Ctcf* loss in the embryonic heart, we examined its genome-wide effects on the transcription. We observe that *Ctcf* is necessary to activate the cardiac developmental transcription program, and it does so by bringing together heart enhancers and target genes. We also assessed the role of CTCF as an insulator in several genomic contexts where binding sites for this factor are located between genes with divergent expression patterns, finding that CTCF does not always act as an insulator and that surely is not the only player responsible to separate distinct regulatory landscapes.

We focused on the IrxA gene cluster that comprises *Irx1*, *Irx2* and *Irx4*, as an example where to understand how 3D genome structure and the regulation of gene expression are mediated by CTCF. To do so, we performed chromosome conformation capture analysis by 4C-seq to interrogate the interaction domains established by the promoters of the genes from the cluster and how they changed upon CTCF deletion. We found that *Ctcf* loss disrupts the 3D genome structure of the cluster by markedly affecting the *Irx4* interaction domain. This had an impact on gene expression of the members of the cluster and its neighboring genes. *Irx4* was downregulated, and on the contrary *Irx1*, *Irx2* and the first three telomeric genes from the cluster, *Ndufs6*, *Mrlp36* and *Lpcat1* were upregulated. Based on the 4C-seq profile we found that the regulatory landscape of *Irx4* is delimited by CTCF binding sites, where different cardiac enhancers are located. Finally, we also found that a specific CTCF binding site located between *Irx2* and *Irx4* is necessary to establish the 3D genome structure of the *Irx4* locus and for proper *Irx4* gene expression.

We show that the 3D structure mediated by the architectural protein CTCF is necessary for proper heart development.

RESUMEN

La célula necesita ejercer un alto control sobre la expresión génica para que la información fluya adecuadamente en tiempo y en espacio. Para lograr esto, la información almacenada en el genoma debe estar altamente organizada para que se despliegue correctamente. Cada vez hay más evidencia sugiriendo que la estructura tridimensional (3D) del genoma juega un papel importante en su organización y función. Diferentes proteínas arquitecturales que se unen a la cromatina son elementos clave en la estructura 3D del genoma, entre estas proteínas CTCF tiene un papel importante en este proceso. Nuestro objetivo era explorar el papel de la estructura 3D mediada por CTCF en el desarrollo de corazón. Encontramos que *Ctcf* es esencial para el desarrollo cardiaco, y que su pérdida da lugar a malformaciones cardiacas que eventualmente conllevan a la muerte embrionaria. Para obtener una visión global del impacto de la pérdida de *Ctcf* en el corazón embrionario, examinamos los efectos en la transcripción de todo el genoma. Observamos que *Ctcf* es necesario para activar el programa transcripcional de desarrollo cardiaco, y lo consigue cuando une enhnacers de corazón y genes blanco. También evaluamos el papel de CTCF como aislador en varios contextos genómicos, en los cuales sitios de unión de este factor están localizados entre genes con patrones de expresión divergente, encontramos que CTCF no siempre actúa como aislador y que seguramente no es el único elemento responsable de separar dominios regulatorios distintos.

Nos enfocamos en el cluster IrxA que contiene los genes *Irx1*, *Irx2* y *Irx4*, como un ejemplo para entender como la estructura 3D del genoma y la regulación de la expresión génica son mediadas por CTCF. Para conseguir esto, realizamos un análisis de captura de la conformación del cromosoma, por 4C-seq, para interrogar los dominios de interacción establecidos por los promotores de los genes del cluster y como estos cambiaban al eliminar CTCF. Encontramos que la pérdida de *Ctcf* desestabiliza la estructura 3D del cluster, afectando marcadamente el dominio de interacción de *Irx4*. Esto tuvo un impacto en la expresión génica de los miembros del cluster y de sus miembros vecinos. *Irx4* estaba downregulado, mientras que *Irx1*, *Irx2* y los tres primeros genes teloméricos del cluster, *Ndufs6*, *Mrlp36* y *Lpcat1* estaban upregulados. Basados en el perfil de 4C-seq, encontramos que el dominio regulatorio de *Irx4* está delimitado por sitios de unión a CTCF, donde se localizan varios enhancers cardiacos. Finalmente, también encontramos que un sitio específico de unión a CTCF localizado entre *Irx2* e *Irx4* es necesario para establecer la estructura genómica 3D del locus de *Irx4*, así como para su expresión génica adecuada.

Aquí mostramos que la estructura 3D mediada por la proteína arquitectural CTCF es necesaria para el desarrollo cardiaco adecuado.

Index

List of figures

List of Tables

List of acronyms

3D	Three dimensional
AV	Atrio ventricular
BS	Binding site
ChiP-seq	Chromatin Inmunoprecipitacion cuopled with sequencing
DEG	Diferencitally expressed genes
FHF	First heart field
GO	Gene ontology
H&E	Hematoxilyn and Eosin
HS	Hipersensitive site
Kb	kilobase
KO	knock out
LADs	Lamina asociated domains
LINE	Long interspread nuclear element
LTR	Long terminal repeat
Mb	megabase
MIR	mammalian-wide interspread repeats
NL	nuclear lamina
PFA	paraformaldehyde
RNA-seq	RNA-sequencing
RT	room temperature
SHF	second heart field
SINE	short interspread nuclear element
TADs	topological associated domains
TF	transcription factor
TSS	transcription start site
ZRS	zone of polarizing activity regulatory sequence

INTRODUCTION

A major challenge in developmental biology is to understand how a single cell gives rise to a complex multicellular organism. All the information necessary to achieve this lies in the genome, an organism´s complete set of DNA including all of its genes. It is clear that two kinds of information exist within the genome, and to understand this better we can divide the genome into coding and non-coding regions. The coding genome includes exons that ultimately translate into proteins. Remarkably, the exon-coding region comprises only 2% of the genome (R. P. Alexander et al. 2010). The remaining 98% of the genome includes non-coding sequences that are required to control gene expression, eg when and where a gene must be expressed or silenced, and this establishment and maintenance of differential patterns of gene expression is key for proper spatial and temporal flow of information from the genome to the cell during development.

In order to exert tight control over gene expression, information must be highly organized. To achieve this, the genome is packed into a compacted chromatin scaffold, and the basic unit is the nucleosome. It consists of DNA wrapped around a protein histone octamer. This configuration helps to package and organize the genome into the nucleus (van Steensel 2011).

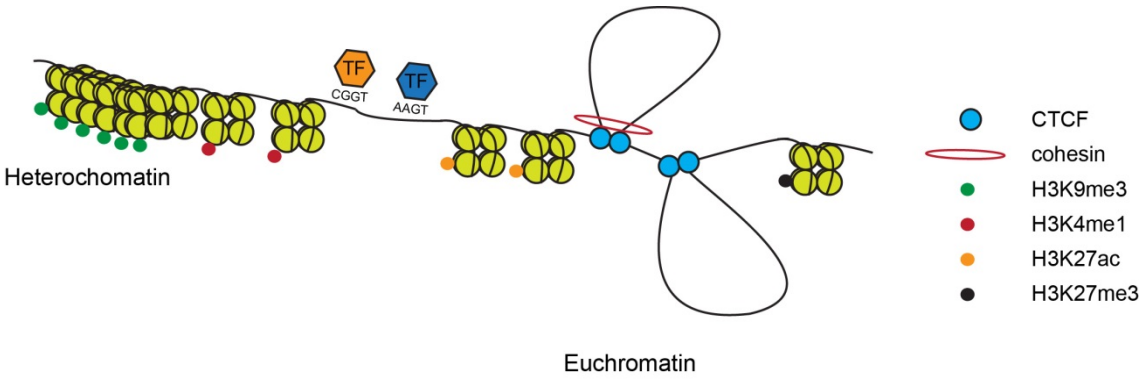

Fig. 1. The regulatory genome and its organization.

The genome is organized in several levels. First, it associates with histone octamers to form the nucleosomes (light green circles). When these nucleosomes are tightly united they form the heterochromatin, a structure that impedes gene expression. The H3K9me3 histone modification is associated with heterochromatin. Nucleosome separation allows the recruitment of several proteins and constitutes the euchromatin. The H3K4me1 histone modification is associated with enhancers that recruit transcription factors to later activate transcription. The H3K27ac modification is present in active regions of the chromatin. Looping occurs to bring together regulatory elements with target promoters, or to group several genes that respond to the same regulatory elements. CTCF (blue circles) and cohesin (red ring) participate in looping stabilization. H3K27me3 associates with repressive chromatin domains.

The functional properties of the genome depend on multiple levels of organization. Structurally, the chromatin is divided into gene-rich, less condensed euchromatin and

gene-poor, highly compacted heterochromatin, usually referred to as active and inactive chromatin, respectively (Solovei et al. 2016). An additional level of organization, the so-called epigenetic phenomenon, is based on modifications to the chromatin, such as methylation and acetylation of specific residues on specific histones, which serve to label regions of DNA as active, inactive, promoters, enhancers, among others. At the level of the primary nucleotide sequence, a myriad of binding sites exist for the diverse repertoire of transcription factors that help to establish tissue identity (Fig. 1). Beyond these well-characterized interactions, increasing evidence suggests that the three-dimensional (3D) genome structure, i.e. the conformation of the chromosomes, also plays a major role in organization and genome function (Bonora et al. 2014).

1. Genomic regulatory elements

The main goal of nuclear organization is to ensure proper gene expression. Genes that require to be expressed depend on different regulatory elements that are distributed along the genome. Some of the main types of regulatory elements are promoters, insulators, repressors and enhancers (Merkenschlager and Odom 2013; Symmons and Spitz 2013). A promoter is a region of DNA that initiates transcription of a given gene. An insulator is a region of DNA that, when placed between two genes, prevents the regulatory elements acting on one gene act over the second gene, and vice versa. Repressors inactivate transcription and are related to the recruitment of the Polycomb complexes, CpG islands or the association with the nuclear lamina (Blackledge et al. 2015). An enhancer is operationally defined as a region of DNA that can activate transcription from a target promoter in an orientation- and location-independent manner (Smallwood and Ren 2013). Enhancers can be physically close to the gene, within the gene body, or can be located many hundreds of kilobases from the gene.

Enhancers act as scaffolds where transcription factors bind to remotely control gene expression (Garcia-Gonzalez et al. 2016). Increasing evidence suggests that remote regulatory elements such as enhancers come into proximity with the target gene by looping, and that this looping can be protein-mediated (Sexton and Cavalli 2015). This looping is not random. There are examples where looping is necessary for proper gene expression (Ribeiro de Almeida et al. 2011), and others where looping seems to poise the genes before transcription (Ghavi-Helm et al. 2014). At the same time, genes also need to be shielded from other regulatory elements that could cause inadequate gene expression (Handoko et al. 2011). This proposed looping mechanism between

enhancers and target promoters is an example of how 3D structure influences gene expression (Fig2).

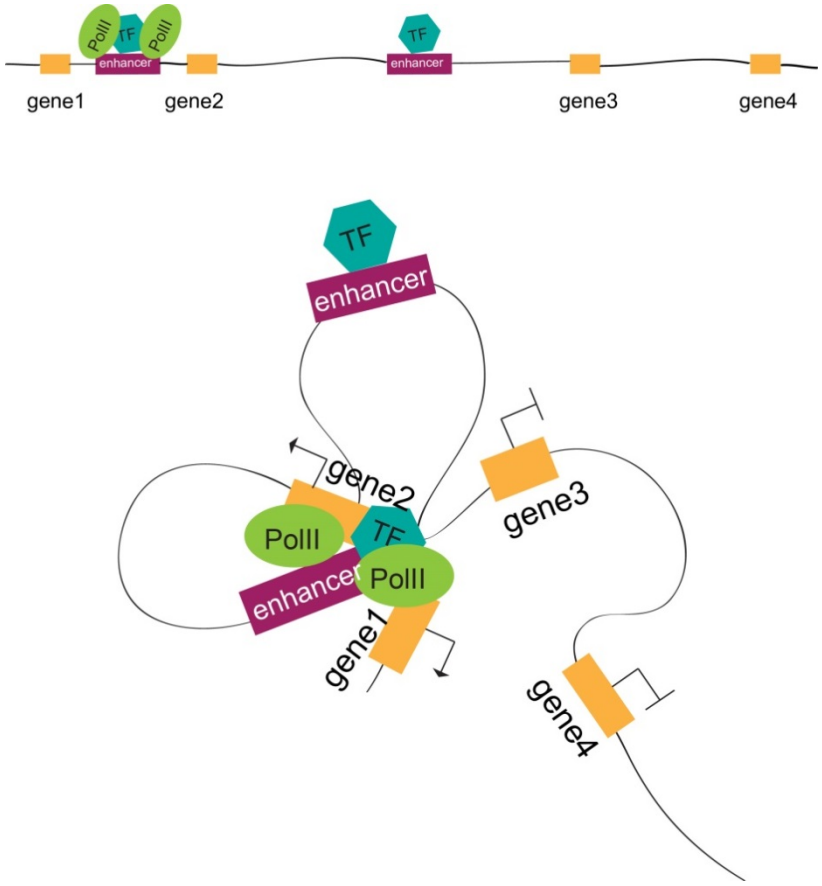

Fig. 2. Enhancer promoter interactions.

Regulatory elements, such as enhancers, can be several kilobases away from their target gene, like the depicted on the top panel. They come into close proximity via a looping mechanism to activate gene expression, like depicted on the bottom panel.

2. The three-dimensional structure of the genome

The three-dimensional **(**3D) genome structure refers to the spatial distribution of the chromatin in the nucleus, and evidence suggests that it is not random (Bickmore 2013; Cremer and Cremer 2001). Rather, it is organized to achieve the most efficient spatial distribution to guarantee proper gene regulation (Bickmore 2013). This is reflected in the nuclear architecture by the existence of structural and functional units, including chromosome territories, lamina-associated domains (LADs) (Bickmore 2013), and topologically associating domains (TADs) (Dixon et al. 2012).

A chromosome territory refers to the space that each chromosome occupies in the nucleus. However, it is not a fixed position. In some cells, the tendency is for gene-rich chromosomes to reside in the middle of the nucleus and for gene-poor chromosomes to reside in the nuclear periphery. In other cells, chromosomes are distributed according to their size; large chromosomes reside in the periphery and smaller ones lie inside the nucleus (Cremer and Cremer 2001).

The nuclear lamina (NL) is a filamentous meshwork composed of lamin proteins located in the nucleus periphery (Kind and van Steensel 2010). Varying sizes (0.1–10 Mb) of DNA associates with the NL, forming LADs. LADs usually contain gene-poor chromatin, and genes within LADs are generally transcriptionally silent (Kind and van Steensel 2010). However, genes associated with LADs dynamically change during development depending on whether they are turned on or off (Peric-Hupkes et al. 2010).

There are many interactions that occur simultaneously inside a chromosome. However, they occur more frequently in discrete domains that have an aproximate size of 1 megabase, indicating that the chromosomes are compartmentalized (Dixon et al. 2012). These discrete domains are the TADS. Interactions between TADs are much less frequent, and they are delimited by the TAD boundaries, which are enriched in CCCTC-binding factor (CTCF) sites, housekeeping genes, and interspersed repeats of the SINE family (Dixon et al. 2012). Inside the TADs occur interactions at lower scales, connecting regulatory elements with their target genes (Dixon et al. 2012; Nora et al. 2013). Accordingly, the current view is that gene expression regulation occurs largely inside a delimited region of DNA (Dixon et al. 2012).

The discovery of TADs was a breakthrough in the genomic architecture field because they shape the way we now understand the genome in 3D. For example, long-range interactions were known before the discovery of TADs; such was the case for *Sonic hedgehog* (*Shh*) in the mouse. The regulation of this gene was striking because an enhancer element necessary for proper *Shh* expression and function was located almost 900 kb distant from the gene (Anderson et al. 2014). This long-range interaction is still considered to be surprisingly long; however, it is now considered as an intra-TAD interaction. Another example is the *HoxD* gene cluster. It has different regulatory elements at both sides of the cluster. Each side, that spans almost 1 Mb, acts upon specific members of the cluster (Lonfat and Duboule 2015). This was better understood after the description of the TADs because each half of the cluster belongs to a different

TAD (Lonfat and Duboule 2015) and, as mentioned above, the interactions between TADs are less frequent.

How the 3D structure is formed and maintained is still an open question. Although, several proteins are involved in maintaining the genomic 3D structure, the most relevant in mammals is CTCF (Ong and Corces 2014).

3. The architectural protein CTCF as a genome organizer

CTCF is an eleven zinc finger DNA-binding protein that was first identified as a transcriptional regulator of the *c-myc* oncogene (Lobanenkov et al. 1990). It is implicated in many processes related to gene regulation including transcription (Vostrov and Quitschke 1997; Vostrov et al. 2002), imprinting (Hark et al. 2000), barrier insulation (Bell et al. 1999), looping and long-range interactions (Handoko et al. 2011), and X chromosome inactivation (Chao et al. 2002). It is highly conserved among eukaryotes, but is not present in yeast, *Caenorhabditis elegans* or plants (Ong and Corces 2014).

Tens of thousands of binding sites for CTCF are found scattered throughout the genome (H. Chen et al. 2012a). A systematic analysis of CTCF binding sites revealed that they can be ubiquitous or cell type-specific (H. Chen et al. 2012a). Several post-translational modifications, such as phosphorylation and poly(ADP-ribosyl)ation (PARylation), can interfere with CTCF binding to chromatin. Moreover, several CTCF binding partners, such as the transcriptional regulator Kaiso and the transcription factor YY1, have been reported (Phillips and Corces 2009).

Initial work using reporter assays with CTCF suggested that it acted as an insulator, by preventing the activation of a promoter by an enhancer when a CTFC binding site was placed between them (Yang and Corces 2012). Genome wide analysis of CTCF binding site locations suggested that this protein establishes domain boundaries in chromatin, separating heterochromatin from euchromatin (H. Chen et al. 2012a), or delimiting chromatin epigenetic states (Merkenschlager and Odom 2013). A study in mouse embryonic stem cells (mESCs) involving chromatin interaction analysis and paired end tagging (ChiA-PET) indicated that CTCF organizes the genome in domains where each region is enriched in a specific combination of histone modifications (Handoko et al. 2011). Following the discovery of TADs, it was understandable to find CTCF in the boundaries of TADs, either alone or in combination with other factors. Strikingly, only 15% of the binding sites were found inside the boundaries (Dixon et al. 2012).

CTCF has been described as the master weaver of the genome (Phillips and Corces 2009), and today it is known as an architectural chromatin-associated protein with a pivotal role in organizing the genome (Ong and Corces 2014) together with cohesin and mediator complexes (Bonora et al. 2014; Kagey et al. 2010; Phillips-Cremins et al. 2013). Several other proteins have also been reported to participate in long-range interactions such as SATB1 (Yasui et al. 2002), among others, which could also be considered as architectural proteins. However, CTCF, cohesin and mediator are currently recognized as the main architectural proteins (Bonora et al. 2014).

Cohesin is a ring-shaped complex that was initially described as mediating sister chromatid cohesion (Nasmyth and Haering 2009), but recently a role in long-range interaction has also been described (Lee and Iyer 2012). Similar to CTCF, cohesin binds genome-wide, and 89% of cohesin binding-sites are co-occupied by CTCF (Lee and Iyer 2012). Also, cohesin binding at certain CTCF binding sites is CTCF-dependent. The most attractive explanation for the combination of these two proteins in mediating long-range interaction is that cohesin maintains the loop by holding together the two sections of DNA. Although many sites are co-occupied, there is still a great proportion of single CTCF or cohesin binding sites, indicating that they can act together and independently of each other. Indeed, some reports describe that cohesin can mediate long-range interaction in a CTCF-independent manner (Lee and Iyer 2012). On the other hand, mediator, a transcriptional coactivator, has recently been described to play a role with the help of cohesion in bringing together enhancers and promoters in ESCs (Bonora et al. 2014; Kagey et al. 2010; Phillips-Cremins et al. 2013). Further studies have demonstrated that while these three proteins are all able to mediate long-range interactions, they do not work at the same time. The most common interacting partners are CTCF/cohesin or mediator/cohesin (Phillips-Cremins et al. 2013). Also, CTCF seems to mediate constitutive and larger interactions, up to 1 Mb (Phillips-Cremins et al. 2013). Conversely, mediator seems to facilitate cell-specific interactions and mediator alone or together with cohesin facilitates looping of 300–600 kb (Phillips-Cremins et al. 2013).

In vitro, the consequences of the absence of cohesin or CTCF in the global 3D structure are different. The absence of both proteins decreases intra-TAD interactions, but only the absence of CTCF increases inter-TAD interactions (Zuin et al. 2014), suggesting that although CTCF alone does not form a TAD boundary, it is necessary to maintain it.

The most notable and recent discovery in CTCF biology is that the orientation of its non-palindromic motif is related to TAD formation and hence to boundary demarcation (Rao et al. 2014). Evidence suggests that divergent CTCF binding sites establish a TAD boundary (Gomez-Marin et al. 2015; Rao et al. 2014), and that convergent binding sites lie inside of the TADs (Rao et al. 2014). This, together with the fact that only 15% of the CTCF binding sites are in the TAD boundaries, strongly suggests that it is not the number of CTCF binding sites that demarcates a boundary, but rather the combination of the CTCF binding site and its orientation. Moreover, recent work has shown that intra-TAD CTCF orientation also affects looping and/or gene expression (de Wit et al. 2015; Guo et al. 2015).

There are several examples where CTCF helps to finely tune gene expression (Ong and Corces 2014; Yang and Corces 2012) by bringing together regulatory elements with target promoters. Such is the case of the protocadherins-alfa (PCDH-α) gene cluster in neurons, where CTCF mediates looping between distinct promoters at the PCDHA cluster and the distant enhancer HS5-1 (Hirayama et al. 2012). CTCF can also influence V(D)J recombination in B-cells by promoting distal over proximal Vk elements usage (Ribeiro de Almeida et al. 2011). Since gene expression is more dynamic than TAD structure (in time and space), it is understandable that the majority of CTCF binding sites lie inside TADs.

The full knock-out of CTCF is embryonic lethal (Fedoriw et al. 2004; Heath et al. 2008; Moore et al. 2012; Wan et al. 2008). Furthermore, conditional deletion in several systems alters proper development of B-cells, T-cells, neurons and macrophages (Heath et al. 2008; Nikolic et al. 2014; Ribeiro de Almeida et al. 2009; Ribeiro de Almeida et al. 2011; Shih et al. 2012), leads to apoptosis in limbs and early nervous system (Soshnikova et al. 2010; Watson et al. 2014) and affects the cell cycle in T-cells (Heath et al. 2008). The role of this protein in heart development is currently unexplored.

4. How to look at the genome in 3D?

Historically, fluorescent *in situ* hybridization (FISH) was the method of choice used to gain insight into the 3D structure of chromatin. FISH consists of fluorescently labeling regions of DNA to visualize them under the microscope (Bickmore 2013). Different sizes of DNA can be labeled, from kilo base-ranged regions up to entire chromosomes. Also, DNA and sub-nuclear compartments can be labeled simultaneously (Ferrai et al.

2010). Labeling DNA and/or nuclear sub-compartments showed that the positioning of genes and chromosomes was not random.

At the chromosome level, FISH showed that chromosomes in the interphase nuclei have a preferred position, the chromosomes territories. This position is not fixed and changes in a cell-specific manner (Bickmore 2013; Ferrai et al. 2010). At the gene level, active genes tend to be in the center of the nucleus and inactive genes tend to be in the periphery. Moreover, genes can be associated with nuclear sub-compartments and these associations are related to gene repression or activation, eg if genes are associated with the NL or RNAPolII, respectively. FISH was also used to visualize long-range chromatin interactions between target genes and regulatory elements located distant from the gene. For example the *Shh* gene and the ZRS enhancer that are separated by ~900 kb, and the *Hoxd13* gene and the global control region that are separated by 180 kb (Bickmore 2013).

In 2002, Dekker and colleagues devised a new PCR-based method to detect physical interactions of chromatin, which was termed 3C, for chromosome conformation capture. This approach consists of chromatin crosslinking (with formaldehyde), restriction enzyme-mediated digestion, and finally ligation of the chromatin with low amounts of DNA ligase to favor intramolecular ligation events, allowing the capture by PCR of regions of the genome in close spatial proximity *in vivo* (Dekker et al. 2002). Basically, the 3C approach answers the question if the region A is interacting with the region B.

With the advent of new technologies, the 3C approach was subsequently modified and coupled with microarrays, and later with massive sequencing (de Wit and de Laat 2012). From the evolution of this method, new 3C-based approaches arose: circularized chromatin conformation (4C- on chip and later 4C-seq), chromatin conformation capture carbon copy (5C), chromatin immunoprecipitation coupled with chromatin conformation capture (ChIP-loop and ChiA-PET) and chromatin conformation capture coupled with next generation Sequencing (Hi-C). 4C-seq searches for all interacting regions that occur with a specific region of interest, whereas 5C searches in an unbiased manner all interactions that occur inside a genomic region up to Mb sizes. Hi-C searches all interactions that occur genome-wide. ChiA-PET also searches all interactions that occur genome-wide, but it focus on those interactions related to a specific chromatin-bound protein (Risca and Greenleaf 2015) (Fig. 3).

The 3C-based methods give insight into the 3D structure inside the chromosome and to a lesser extent between chromosomes. FISH can also give insight into the 3D structure inside of the chromosome but at a lower resolution. FISH is visually compelling but generally limited to looking at the locations of a few specific targets in a few hundred cells. Both methodologies aim to explore the 3D structure to further understand how it translates into a phenotype, only at different scales and with different resolution. This is why FISH can complement 3C-based methods and vice versa (Bickmore 2013). In this work, we have made extensive use of the 4C-seq approach.

5. Circularized Chromatin Conformation Capture: 4C-seq

The original 3C method is limited by the number of primer pairs designed, and only two sites can be interrogated at the same time. These disadvantages can be overcome with 4C-seq because the massive sequencing step allows the interrogation of all the interactions occurring with a specific region (the viewpoint) at once (de Wit and de Laat 2012).

Several works have proven that 4C-seq is a powerful method to gain insight into genome organization and function. For example, exploring the interaction profiles from housekeeping and cell-specific genes showed that each type of gene presents constitutive and cell-specific interactions, respectively (de Wit and de Laat 2012). Further work showed that inactive genomic regions in ESCs form few interactions. These inactive regions do form interactions in differentiated cells, such as murine embryonic fibroblasts (MEFs). This indicates that the inactive chromatin in ESCs is disorganized and that it becomes organized by forming interactions upon differentiation. Strikingly when a MEF is reprogrammed, the interactions of the inactive regions formed upon differentiation are lost (de Wit et al. 2013). It has also been shown that a genomic site that contains binding sites for several pluripotent factors tends to interact more with similar regions, i.e. those also containing pluripotency factor binding sites. A similar scenario occurs with Polycomb repressed regions (de Wit et al. 2013). In general, genomic regions tend to interact with other genomic regions with similar features. Furthermore, the sequential activation of the *Hoxd* cluster genes coincides with the sequential formation of interactions between the cluster genes and their regulatory elements (Noordermeer et al. 2011). Several works support the idea that enhancers can interact with promoters without activating their transcription, thereby gaining a poised regulatory architecture (de Wit and de Laat 2012; Ghavi-Helm et al. 2014). Given that 4C-seq reveals, in an unbiased manner, all the interacting regions, it

can therefore be used to search previously unknown regulatory regions (van de Werken et al. 2012b), allowing a better understanding of the regulation of genes (van Weerd et al. 2014) and even explaining complex disease-related phenotypes (Smemo et al. 2014).

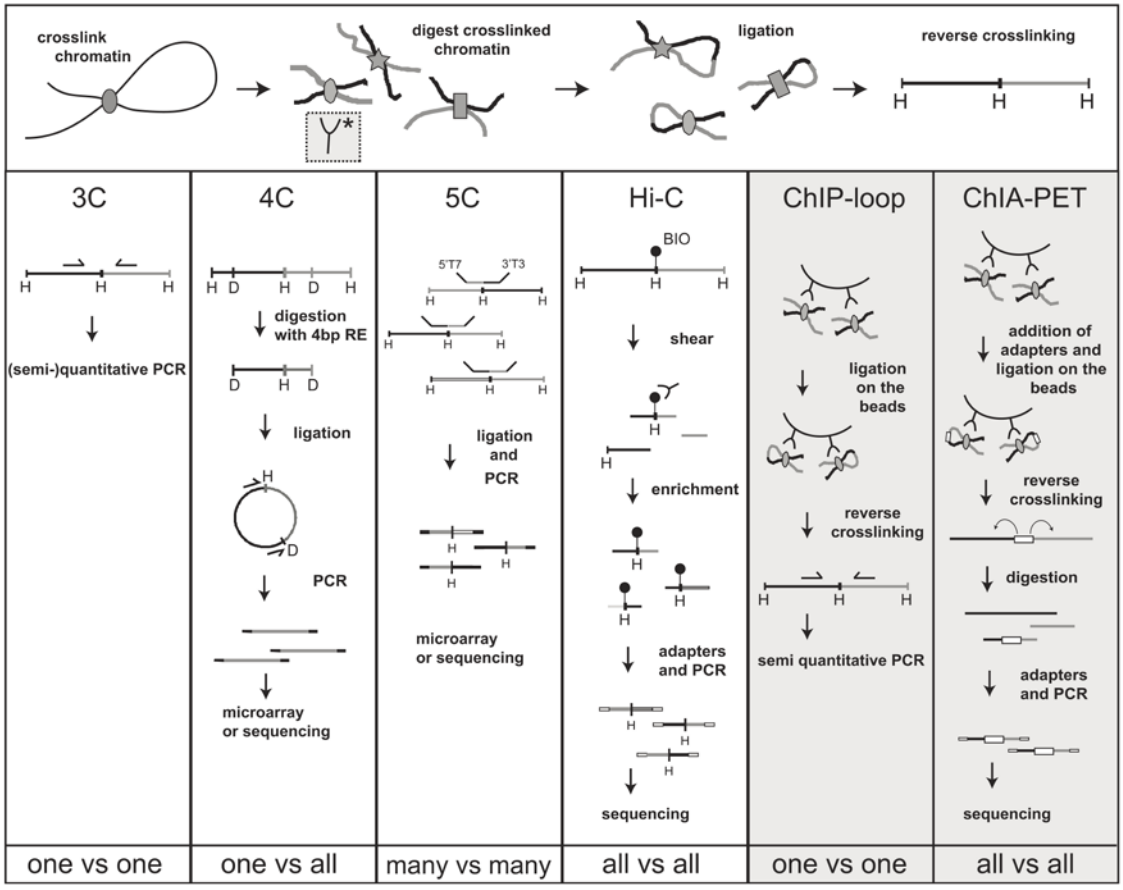

Fig. 3. Overview of 3C-based methods.

The 3C-based methods aim to capture interactions occurring in the crosslinked chromatin. What varies between methods is the number of genomic regions interrogated (3C and 4C focus on one region, 5C focuses on many and Hi-C interrogates the whole genome) and the size of the genomic region explored (3C explores only one region, 4C explores the whole genome, 5C explores up to 1Mb, and Hi-C explores the whole genome). The top panel depicts the steps common to all methods. The middle panels depict the variations for each method. The bottom panels illustrate a descriptive summary of each method. Taken from (de Wit and de Laat 2012).

6. Testing the functional genome

Studies on 3D structure have shown that physical interactions translate into phenotypes and that constant crosstalk occurs between coding and non-coding sequences, which is necessary for gene regulation. To further understand the relevance of the coding and non-coding sequences and how they work, several

strategies have been devised to address these issues, both *in vitro* and *in vivo*. Some of these are common to both types of sequences and others are sequence type-specific. Coding sequences can be deleted, overexpressed and mutated. The function of non-coding sequences can be tested by the ability of a region of DNA to activate expression of a reporter gene, e.g. transient transgenic assays, or to repress expression. Non-coding sequences can also be deleted, mutated or inverted. Genetic manipulation of the DNA is required for these processes and can be time consuming, particularly if the goal is to test the non-coding sequence *in vivo* by generating transgenic organisms. The CRISPR/Cas9 technology has proven to be a powerful genome-editing tool because of its efficiency (Seruggia and Montoliu 2014).

6.1 Genome editing by CRISPR/Cas9

The CRISPR/Cas9 system has been used successfully in a wide variety of fields (Seruggia and Montoliu 2014). Despite its early discovery as a bacterial genome-editing tool (Jinek et al. 2012), this technology has proven beneficial to get insight into mammalian genome organization and function. Regarding 3D structure, CRISPR/Cas9 editing has been used to disrupt boundary elements that are required to establish functional chromatin domains at the Hox clusters (Narendra et al. 2015), and to prevent aberrant interactions that could lead to cancer (Flavahan et al. 2016). In *C. elegans*, boundary disruption showed that the partitioning of the X chromosome resembles the TADs in mammals (Crane et al. 2015). A seminal work with the CRIPSR/Cas9 technology was the demonstration that a TAD disruption *in vivo* leads to pathogenic gene-enhancer interaction, which recapitulated limb malformations (Lupianez et al. 2015). Given that CTCF binding polarity (binding site orientation) is related to TAD boundary formation and to intra-TAD interactions (Gomez-Marin et al. 2015; Rao et al. 2014), CRISPR/Cas9 technology was used to invert CTCF binding sites. Strikingly, this inversion disrupted looping (de Wit et al. 2015) and altered enhancer-promoter interactions necessary for proper gene expression (Guo et al. 2015).

7. Heart development

In mammals, the heart is the first organ to form during embryogenesis. After gastrulation, the primitive streak arises in the bilaminar disc. Primitive streak cells migrate anteriorly to form the cardiac precursor at embryonic day (E) 6.5. Cardiac progenitors then spread towards the two sides of the midline to form the cardiac crescent, initiating myocardial differentiation at E7.5. Both sides of the cardiac crescent move towards the midline and fuse, giving rise to the primitive cardiac tube (E8.0) with

arterial and venous poles. The cardiac progenitors that give rise to the primitive cardiac tube are a cell population called the first heart field (FHF) that will contribute to the left ventricle, the atrioventricular canal and both atria (Buckingham et al. 2005). The resulting cardiac tube continues growing by the addition of pharyngeal mesoderm cells at both poles. This cell population is termed the second heart field (SHF), and will contribute to the outflow tract and all other heart regions except the left ventricle (Buckingham et al. 2005). Simultaneously, the heart tube undergoes rightward looping and subsequently divides into chambers, one common atrium and one common ventricle, connected by the atrioventricular canal. The venous and arterial pole of the heart can be distinguished as the cardiac inflow and outflow tract, respectively. At E9.5, cells coming from the proepicardium, an extracardiac cell population cover the myocardium to form the epicardium. The epicardial layer is involved in the formation of a subset of coronary endothelial cells, coronary smooth muscle cells, part of the AV valves, and cardiac fibroblasts (Luxan et al. 2016). By E10.5, two types of myocardium coexist: the proliferative compact myocardium and the trabecular myocardium. At this stage, cardiac septation begins to physically separate the heart chambers. In the atrioventricular canal and outflow tract, a transcriptional program that blocks myocardium identity is activated, allowing valve formation to occur. Trabecular remodeling begins after formation of the four chambers. At E12.5, cardiac neural crest cells invade the outflow tract and participate in the proper development of the outflow tract valves (Luxan et al. 2016). The heart is fully functional by E14.5, and supports systemic and pulmonary blood circulation through the aorta and pulmonary trunk, respectively (Buckingham et al. 2005) (Fig. 5).

The cardiac transcriptional program has been extensively characterized (Bruneau 2002). Consequently, many key genes involved in different aspects of heart development have been described, including *Nkx2*.5, *Hand1*, *Mef2, Islet* (Buckingham et al. 2005), *Irx4*, *Tbx5* (Bruneau 2002)*, Notch* (MacGrogan et al. 2010)*, Pitx2* (Franco and Campione 2003), *Nppa* (Bruneau 2002), *Hopx* (Trivedi et al. 2010), etc. Among these transcription factors, *Nkx2*.5 plays a key role in the differentiation and morphogenesis of the early developing heart (Bruneau 2002), and is expressed in the two distinct heart fields. For this reason, to assess the relevance of the absence or overexpression of a given gene using the Cre/LoxP system, Cre recombinase is typically driven by the *Nkx2.5* promoter (Stanley et al. 2002).

In addition to the study of genes directly related to heart morphogenesis and patterning, some works have addressed the relevance of epigenetics in heart development. These works have focused on several Polycomb members (J. M.

Alexander et al. 2015; Beketaev et al. 2015; L. Chen et al. 2012b) and chromatin remodeling factors (Lickert et al. 2004; Luna-Zurita and Bruneau 2013; Takeuchi et al. 2011). The study of the three dimensional structure in heart development remains largely unexplored (van Weerd et al. 2014).

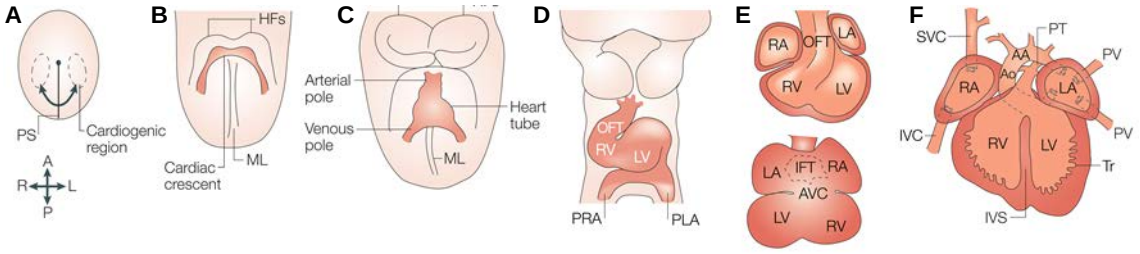

Fig. 4. An overview of heart development.

A. In the mouse, cardiac progenitor cells originate in the primitive streak (PS) at E6.5. **B**. At E7.5, these groups of cells migrate towards the two sides of the midline (ML) and form the cardiac crescent. **C**. At E8.0, both sides of the cardiac crescent will fuse in the midline and form the cardiac tube. **D**. At E8.5, the cardiac tube undergoes rightward looping and divides into chambers. **E**. At E10.5, the four chambers are defined but are not yet fully separated (upper panel is a frontal view, lower panel is a dorsal view). **F**. At E14.5, the four chambers are separated and the heart is fully functional. HFs = head folds; OFT = outflow tract; RV = right ventricle; LV = left ventricle; PRA = primitive right atrium; PLA = primitive left atrium; RA = right atrium; LA = left atrium; IFT = inflow tract; AVC = atrioventricular canal; SVC = superior vena cava; IVC = inferior vena cava; IVS = interventricular septum; Ao = aorta; AA =aortic arch; PT = pulmonary trunk; PV = pulmonary vein; Tr = trabeculae. Taken from (Buckingham et al. 2005).

7.1. *Iroquois* genes

The *iroquois* or *Irx* are a group of genes coding for homeodomain transcription factors with multiple roles in early development. *Irx* genes are expressed in the developing heart. Some members of this group have been shown to have roles in heart morphogenesis (Kim et al. 2012).

Iroquois genes were first identified and described in *Drosophila* as playing a role in neural development (Gomez-Skarmeta and Modolell 2002). These genes were named after the Iroquois American Indian tribe because their absence in *Drosophila* produced flies without lateral bristles in the thorax, and this phenotype resembled the hair style of the American tribe (Cavodeassi et al. 2001).

In *Drosophila,* there are three iroquois genes located in one cluster (Cavodeassi et al. 2001). The molecular characterization of these genes in the fly allowed their subsequent identification in vertebrates. In mammals, six members of the *Iroquois* family were identified(Bosse et al. 1997; Houweling et al. 2001). Mapping analysis of the six *iroquois* genes revealed that they are distributed in two clusters (Cluster A and B). Each cluster, as in the fly, contains three genes. Cluster A comprises *Irx1*, *Irx2* and

Irx4, and Cluster B comprises *Irx3*, *Irx5* and *Irx6*. Strikingly, each cluster spans a large genomic region of ~1.5 Mb (Cluster A) and ~1 Mb (Cluster B). Since each cluster only harbors three members, the *iroquois* genes constitute a good example of what is called "gene deserts" (Gomez-Skarmeta and Modolell 2002). In the fly, the *iroquois* cluster also occupies a large genomic territory of 130 kb (Gomez-Skarmeta and Modolell 2002). Sequence comparison between the six genes showed that each member of Cluster A has a paralogous gene in Cluster B (*Irx1~Irx3, Irx2~Irx5, Irx4~Irx6*), and it was suggested that the two clusters came from an ancestral duplication of a three member cluster (Peters et al. 2000). Moreover, the genomic organization of each cluster, which includes the sequential order of the genes inside the cluster, the orientation in which they are transcribed and the coding and non-coding sequences, is conserved in mammals (Gomez-Skarmeta and Modolell 2002) (Fig. 5).

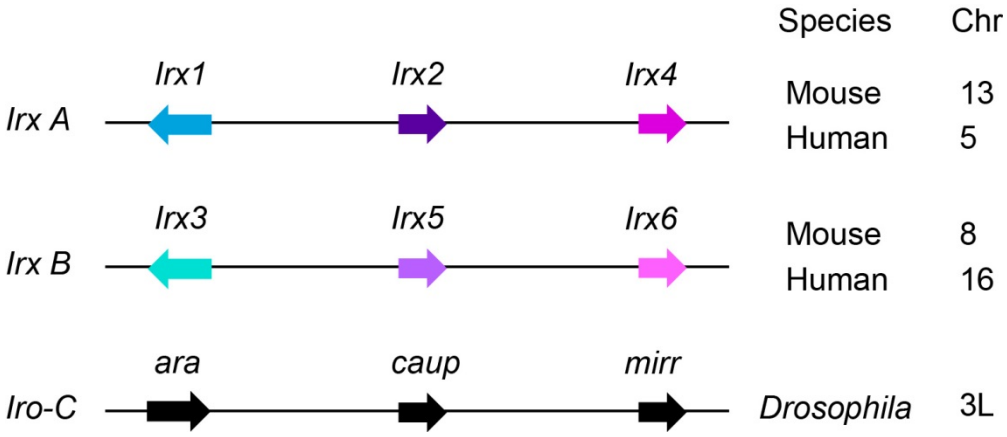

Fig. 5. Genomic organization of *iroquois* genes.

Humans and mouse have six *iroquois* genes distributed in two clusters. Each cluster spans ~ 1 Mb and contains only three genes each. Each gene from the cluster A has a paralogous member in the cluster B. Paralogous genes are depicted in the same colour, darker in the cluster A and lighter in the cluster B. The number of genes in each cluster, their location and distribution inside the cluster, together with the orientation of transcription, the coding sequences and the non-coding sequences constitute the genomic organization of these genes, and is conserved in mouse and humans. In the fly, there are only three *iroquois* genes and they also span a large territory (~130 kb). Furthermore, the orientation is not the same as in mammals for all the genes. Modified from (Gomez-Skarmeta and Modolell 2002).

Initial studies of these genes in the mouse focused on their expression pattern. The aim was to determine whether they played a similar role as in *Drosophila* (Bosse et al. 1997). These studies showed that the *iroquois* genes were expressed not only in the developing nervous system, but also in the limb, in the developing heart and in other territories (Bosse et al. 1997; Christoffels et al. 2000; Houweling et al. 2001).

The expression patterns of *Irx1*, *Irx2*, *Irx3* and *Irx5* in the developing mouse are highly similar and their expression territories are wide (Houweling et al. 2001). By contrast,

Irx4 and *Irx6* show a divergent and more restricted expression pattern (Houweling et al. 2001). In the nervous system, *Irx1*, *Irx2*, *Irx3* and *Irx5* expressed in the median plane of the neural tube and the otic cap but at different levels. There are other nervous system territories where not all of the four genes are present, like the ganglion of the eight cranial nerve (Houweling et al. 2001). The non-overlapping expression patterns are better appreciated in limbs and in the developing heart. In the limbs at E10.5, *Irx1* is expressed in the apical ectodermal ridge of the forelimb, *Irx2* is expressed in all the limb bud and *Irx3* presents a gradient expression in the proximodorsal margin of the limb (Bosse et al. 1997). In the developing heart, all six *iroquois* genes are expressed. From E9.5 to E12.5 *Irx1*/*Irx2* expression is restricted to the interventricular septum, *Irx4* is broadly expressed in the ventricles, *Irx3* expression is restricted to the trabecular myocardium, and *Irx5*/*Irx6* are expressed in atrial and ventricular endocardium (Kim et al. 2012) (Fig. 6).

A functional role for these genes in neural and cardiac development has been shown (Bao et al. 1999; Cavodeassi et al. 2001; Gaborit et al. 2012). In the context of this project, the developing heart, *Irx3* and *Irx5* regulate atrioventricular canal morphogenesis and outflow tract formation (Gaborit et al. 2012). *Irx4*, the member of the family with the broadest expression pattern in the heart, establishes ventricular identity (Bao et al. 1999). *Irx1–5* are expressed in the adult heart and they have also been studied in this context. The expression territories of the *iroquois* genes expressed in the adult heart, with the exception of *Irx3*, are similar to those in the developing heart (Kim et al. 2012) (Fig. 6). In the adult heart, *Irx3* participates in the ventricular conduction system (Zhang et al. 2011), *Irx5* regulates the cardiac repolarization gradient (Costantini et al. 2005), and *Irx4* is necessary to prevent abnormal ventricular gene expression that can later lead to cardiomyopathy (Bruneau et al. 2001).

Given that the *iroquois* genes share expression territories, especially in the nervous system, it was hypothesized that these genes shared regulatory elements such as enhancers. A study carried out in *Xenopus* and zebrafish strongly supported this idea, since a functional screening of enhancers inside the non-coding regions of Cluster B showed that several enhancers were capable of driving expression in shared *iroquois* expression territories (de la Calle-Mustienes et al. 2005). A similar scenario was found for Cluster A (Tena et al. 2011). Also, since the genomic organization is conserved across species, it was hypothesized that their genomic organization was necessary for proper *iroquois* gene regulation possibly by sharing enhancers located inside conserved non-coding regions (Cavodeassi et al. 2001). Further work using 3C

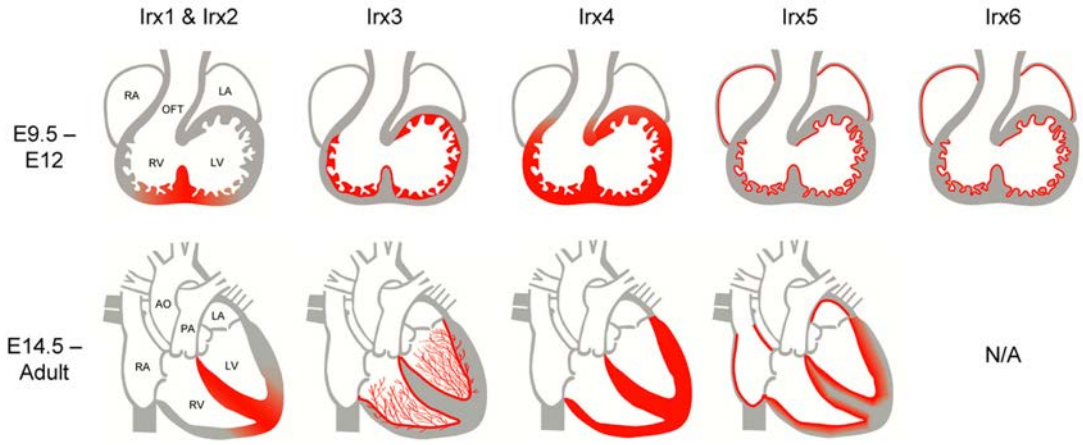

Fig. 6. *Iroquois* gene expression in the heart.

All six *iroquois* genes are expressed in the heart. The top panel shows the expression patterns in the early mouse developing heart. *Irx1* and *Irx2* are expressed both in the interventricular septum and in the near compact and trabecular myocardium. *Irx3* is only expressed in the trabecular myocardium. *Irx4* is highly expressed in ventricle myocardium. *Irx5* and *Irx6* are expressed in all the endocardium lining the atria and ventricles. The bottom panel shows the expression patterns from E14.5 to adult stage. Taken from (Kim et al. 2012)

methodology (that tests the physical interaction between two regions of DNA) showed that indeed this was the case. In the Cluster A, there is physical interaction between enhancers and promoters and some enhancers are shared by more than one member of the Cluster. Also, the physical interaction between *Irx1* and *Irx2* promoters is conserved in vertebrates and is tissue independent, whereas the interactions between enhancers and promoters are tissue specific. CTCF is present in the promoters of the *iroquois* genes and downregulation of CTCF in zebrafish with morpholinos disrupted *Irx1* and *Irx2* promoter interaction. CTCF downregulation also dysregulated Cluster B gene expression. Thus, this conserved 3D structure seems to be CTCF dependent (Tena et al. 2011). Taking together, the *iroquois* genes seem to be an ideal system to study the relationship between 3D structure and gene expression.

In this work we aimed to understand how the 3D genome architecture, mediated by CTCF, impacts gene expression in the developing mammalian heart.

OBJECTIVES

All the information necessary for proper development and tissue homeostasis lies in the genome. Moreover, gene expression control is necessary for the adequate flow of the information in time and space. The 3D structure of the genome has a role in gene regulation, and the architectural proteins are key players involved in 3D structure. In vertebrates, CTCF is an essential architectural protein necessary for development, whose role in the heart development was unexplored. In this work we aimed to understand how the 3D genome architecture, mediated by CTCF, impacts gene expression in the developing mammalian heart. To achieve this goal, we defined the following objectives:

- Analyze the impact of *Ctcf* deletion in the mammalian developing heart.

- Determine whether 3D structure is altered either globally and/or locally upon *Ctcf* deletion, using 4C-seq.

- Determine whether 3D structure mediated by CTCF controls gene expression.

MATERIALS AND METHODS

1. Mouse strains

1.1. Breeding and genotyping

We used the Cre/LoxP system to delete *Ctcf* in the embryonic heart by crossing two different mouse strains. A mouse line carrying the Cre recombinase construct under the control of the *Nkx2*.5 promoter, which is expressed in myocardium and endocardium (Stanley et al. 2002), was kindly provided by Dr José Luis de la Pompa (CNIC). The mouse line carrying the *Ctcf* floxed allele was generated in the laboratory of Niels Galjart (Erasmus MC, Rotterdam, The Netherlands) (Fig. 7) (Heath et al. 2008).

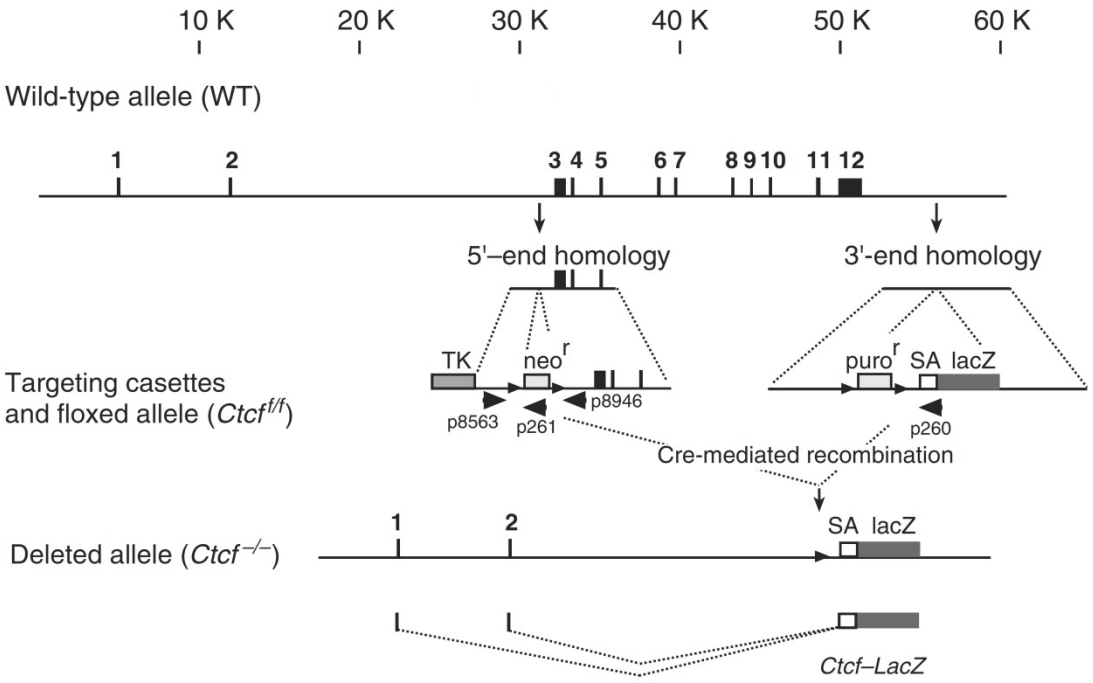

Fig. 7. Conditional targeting of the *Ctcf* gene.

A diagram of the *Ctcf* gene showing the positions of its 12 exons is shown at the top. The DNA length scale is shown in kilobases (K). The resulting *Ctcf* deleted allele is shown at the bottom. Modified from (Heath et al. 2008).

Nkx2.5-Cre; Ctcf$^{fl/+}$ males were crossed with *Ctcf*$^{fl/fl}$ females, giving the following four possible genotypes:

Ctcf $^{fl/fl}$***; Nkx2.5-Cre,*** referred to here also as *Ctcf* mutant or *Ctcf* KO.

Ctcf $^{fl/+}$***; Nkx2.5-Cre***, referred to here also as heterozygous

Ctcf $^{fl/fl}$*,* referred to here also as control

Ctcf $^{fl/+,}$ referred to here also as control.

Table 1. Genotyping primers for *Ctcf* and *Nkx2.5-Cre* alleles

Primer	Sequence	Notes
NKX2.5-Cre	GATGACTCTGGTCAGAGATACCTG	Forward for *Nkx2.5-Cre* allele
NKX-S	GCCCTGTCCCTCAGATTTCACACC	Forward for *Nkx2.5* wildtype allele
Nkx-AS	GCGCACTCACTTTAATGGGAAGAG	Reverse common primer for *Nkx2.5*
p8563	CTAGGAGTGTAGTTCAGTGAGGCC	Forward common primer for *Ctcf*
p8946	GCTCTAAAGAAGGTTGTGAGTTC	Reverse for *Ctcf* wildtype allele
p261	CGGCATCAGAGCAGCCGATTG	Reverse for *Ctcf* floxed allele
p260	TGTCACTAATCTCCACCTCACAG	Reverse for *Ctcf* deleted allele

Primers for the Nkx2.5 allele were taken from (Stanley et al. 2002). Primers for the *Ctcf* allele were taken from (Heath et al. 2008)

For genotyping, each transgene was tested independently. The *Nkx2.5-Cre* allele was detected by the presence/absence of a PCR product using three primers in the same reaction: a common reverse primer and two different forward primers. The wild-type allele PCR product was 200 base pairs (bp) and the *Nkx2.5-Cre* allele was 500 bp. *Ctcf* wild type, floxed, or deleted alleles were detected using three primers in the same reaction: a common forward primer and two different reverse primers. The *Ctcf* wildtype allele PCR product was 383 bp, the *Ctcf* floxed allele was 747 bp, and the *Ctcf* deleted allele was 750 bp. The primers used for this analysis are listed in Table 1 and the PCR conditions for each transgene are shown in Table 2.

Table 2. PCR conditions for the *Ctcf* and *Nkx2.5-Cre* allele

***Ctcf* alleles**			
Initial Denaturation	95ºC	5 min	
Denaturation	94ºC	1 min	35 cycles
Annealing 1	69ºC	1 min	
Annealing 2	60ºC	1 min	
Elongation	72ºC	1 min	
Final Elongation	72ºC	10 min	
Hold	4ºC	∞	
***Nkx2.5-Cre* allele**			
Initial Denaturation	94ºC	5 min	
Denaturation	94ºC	30 sec	30 cycles
Annealing	60ºC	30 sec	
Elongation	72ºC	1 min	
Final Elongation	72ºC	10 min	
Hold	4ºC	∞	

1.2. Embryo collection

Mouse embryos of different genotypes for different experimental procedures were collected at several stages, from E9.5 to E14.5. For transient transgenic experiments, embryos were collected from E9.5 to E13.5.

2. Histology and tissue staining

2.1. Embryo processing

Embryos to be sectioned were dissected in cold phosphate buffered saline (PBS), fixed in 4% PFA overnight at 4°C, dehydrated through a graded ethanol series (30%, 50%, 70%, 90% and 100%; 20 min each) and stored at −20°C until processing. Ethanol was diluted with RNAse-free water. Embryos were rehydrated by reversing the ethanol series and were processed for histological analysis of hematoxylin and eosin staining, RNA *in situ* hybridization on sections, or immunostaining. After rehydration, embryos were washed in xylene for 2×30 min, soaked in xylene:paraffin (1:1) for 30 min at 65ºC and then embedded in paraffin at 65ºC for 2×30 min. Before the paraffin became solid, the embryo inside the paraffin block was disposed vertically under a dissecting microscope to obtain transversal sections of the heart. Once the paraffin was solid, the blocks were stored at 4ºC until further processing.

For RNA *in situ* hybridization of whole-mount embryos, embryos were dehydrated through a methanol series (50%, 75% and 100%; 20 min each) and embryos were stored at -20ºC. Methanol was diluted in PBT (PBS + 1%Triton X-100).

2.2. Hematoxylin and Eosin staining

Sections were incubated at 65ºC for 20 min, dewaxed with xylene for 2×10 min, and rehydrated into distilled water in a decreasing ethanol series (twice with 100%, 90%, 70%, 50%, 30% and finally distilled water; 5 min each; the ethanol was diluted in distilled water). Sections were incubated with Harris Hematoxylin (Merck, reference 1.00317.1000) for 10 min. Then, Hematoxylin was removed with rinsing water. Sections were treated with 1% acid alcohol (1% HCl diluted in 100% ethanol) for 2-5 min. Sections were washed with rinsing water for 5 min. Sections were incubated in aqueous eosin (Merck, reference 1.00317.1000) for 10 min. Then, eosin was removed with rinsing water. Sections were dehydrated through an ethanol series followed by two washes with xylene and mounted with Dpx Mountant medium (Sigma 06522).

2.3. RNA *in situ* hybridization on paraffin sections.

RNA *in situ* hybridization was performed on 7-μm-thick sections following a previously described protocol (Jostarndt et al. 1994; Kanzler et al. 1998) with a few modifications.

On the day prior to *in situ* analysis, sections were incubated at 37ºC overnight. On the next day, they were incubated at 65ºC for 20 min, dewaxed with xylene for 2×10 min, and rehydrated into PBS in a decreasing ethanol series (twice with 100%, 90%, 70%, 50%, 30% and finally PBS; 5 min each). Ethanol was diluted with RNAse-free water. Sections were re-fixed with 4% PFA for 20 min at room temperature (RT) and washed twice with PBS for 5 min. Then, they were incubated 10 min at 37ºC with proteinase K at 10 µg/mL and washed with PBS for 5 min. They were re-fixed with 4% PFA for 5 min at RT and washed with PBS for 5 min. Then, sections were treated with 0.7 N HCl for 15 min at RT while agitating, and washed with PBS twice for 5 min, then were incubated in 0.25% acetic anhydride (acetic anhydride was diluted in 0.1 M triethanolamine at pH 8) for 10 min at RT while agitating, and washed with PBS for 5 min. Sections were given a final wash with RNAse-free water for 5 min before pre-hybridization (see appendix A) for 2 h at 65ºC. Probes were hybridized overnight at 65ºC with pre-hybridization buffer. PBS used for these steps was sterile. The next day, sections were washed twice with post-hybridization buffer I at 65ºC and then twice with post-hybridization buffer II at 65ºC. Later, were washed with maleic buffer (MABT) three times for 5 min each. Then, were blocked for 2 h at RT with blocking solution and then incubated with an anti-DIG antibody 1:2000 (Roche, reference 11 093 274 910) overnight at 4ºC in a humidified chamber. The next day, sections were washed twice with MABT for 10 min while agitating at RT and then washed again with MABT three times for 1 hour while agitating at RT. Then, they were washed with AP-buffer three times for 10 min while agitating at RT and finally developed with BM-purple (Roche, reference 11442074001). After this, sections were washed with PBS twice for 5 min, re-fixed with 4% PFA and washed again with PBS twice for 5 min. They were dehydrated through an ethanol series followed by two washes with xylene and mounted with Dpx Mountant medium (Sigma 06522).

2.4. RNA *in situ* hybridization in whole-mount embryos

Whole mount *in situ* hybridization was performed as previously described (Wilkinson and Nieto 1993). Whole-mount embryos were rehydrated through a decreasing methanol series (75%, 50%, 10 min each) followed by treatment with 6% H_2O_2 for 1 h at RT in the dark. After washing with PBT three times for 10 min, embryos were treated with proteinase K (10 µg/mL) for 7 min (E9.5 embryos) or 12 min (E10.5 embryos) at RT and quickly rinsed with PBT twice and then washed with PBT for 5 min at RT. Embryos were re-fixed for 20 min at RT with 0.25% glutaraldehyde diluted in 4% PFA, followed by one quick rinse with PBT and a further two washes with PBT for 5 min at RT. Embryos were then washed with pre-hybridization buffer at 65ºC for 5 min and

incubated for 2 h with pre-hybridization buffer at 65ºC. Probes were hybridized overnight at 65ºC. On the next day, embryos were washed twice with post-hybridization buffer I at 65ºC and twice with post-hybridization buffer II at 65ºC. Later, embryos were rinsed with TBST and then washed three times for 10 min at RT. Embryos were incubated with blocking solution for 2 h at RT, followed by overnight incubation at 4ºC with anti-DIG antibody 1:2000. The next day, embryos were washed with TBST three times for 5 min, and then washed again with TBST five times for 1 h. Embryos were left overnight in TBST at 4ºC. The next day, embryos were washed three times with NTMT buffer (appendix A) for 10 min at RT and developed with BM-purple. After developing, embryos were washed twice with NTMT buffer and stored in 4%PFA at 4ºC. All treatments and incubations were done with agitation.

Table 3. Primers used for PCR-generated probes

Primer	Sequence
*Tnni2*_F	ACGTGGCTGAAGAGGAGAAA
*Tnni2*_R	TATTGGAGCGAGGCCAAGTA
*Tnni3*_F	GAGACCTCCAAGGTCACCAG
*Tnni3*_R	TTAAACTTGCCACGGAGGTC
*Tnnt1*_F	TGCACTAAAAGACCGCATTG
*Tnnt1*_R	TGGGGGCACTTTATTTTGAG
*Tnnt3*_F	GAAACCAAGACCCAAACTTAC
*Tnnt3*_R	TTTATTCCTAGACCCCAGAAG
*Ndufs6*_F	GGTTTCGGGGTTCAAGTGT
*Ndufs6*_R	TGGTGGGAGCATCCTTTATT
*Hopx*_F	AGCAGACGCAGAAATGGTTT
*Hopx*_R	CCCCTGCCTGTTCTGTTATC
Mrlp36_F	GCCCGAAGTTGATTAAGGAC
Mrlp36_R	CAAGAGCTTTTTGGCTCAGG
Fgf13_F	TCGCTCATCCGGCAAAAGAG
Fgf13_R	GGTTCTGTTATAGAGCCCTCG
SP6 promoter	ATTTAGGTGACACTATAGAA
T7 promoter	GTAATACGACTCACTATAGGG

F stands for forward primer and R stands for reverse primer

2.5. Digoxygenin-labeled riboprobe synthesis for *in situ* hybridazation

Riboprobes used for *in situ* hybridization were either generated by PCR and then transcribed, or directly transcribed from plasmids containing the requisite cDNA. In the former, total RNA from mouse embryos at different stages (depending on the gene of interest) was extracted using the RNeasy Mini kit (Qiagen, reference 74106) and cDNA was synthesized with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, reference 4368814). Primers were designed at the 3´UTR of the gene of

interest. Sp6 and T7 promoter sequences were added at the 5´ end of the forward and reverse primer, respectively. The sense probe was transcribed with SP6 and the anti-sense probe was transcribed with T7. For plasmid probes, plasmids were linearized and transcribed with the appropriate restriction enzyme. Digoxygenin-labeled riboprobe synthesis was performed according to the manufacturer's protocol (Roche). Further details of the probes are shown in Tables 3 and 4.

Table 4. Probes contained in plasmids

Probe	Restriction Enzyme	RNA Polymerase	Origin
Irx1	*XbaI*	T3	José Luis Gómez Skarmeta
Irx2	*BamHI*	T3	José Luis Gómez Skarmeta
Irx4	*NotI, SpeI*	T7	José Luis Gómez Skarmeta
Irx3	*HindIII*	T7	José Luis Gómez Skarmeta
Irx5	*XbaI*	T3	José Luis Gómez Skarmeta
Irx6	*EcoRI*	T7	José Luis Gómez Skarmeta
Nkx2.5	*XbaI*	T7	José Luis de la Pompa
Tbx20	*SacII, NcoI*	Sp6	José Luis de la Pompa
Nppa	*BamHI*	T7	José Luis de la Pompa
Pitix2	*SacII*	T3	José Luis de la Pompa
Tbx2	*EcoRI*	T3	José Luis de la Pompa
Tbx3	*PstI*	T3	Robert Kelly
Tbx4	*XbaI*	Sp6	María A. Ros
Tbx5	*EcoRV*	T7	José Luis de la Pompa

2.6. Immunohistochemistry

For CTCF detection, slides were dewaxed and rehydrated, washed in distilled water and subjected to heat-induced epitope retrieval in citrate buffer. Briefly, 10 mM citric acid pH 6.0 was microwave heated for 3 min, then slides were microwave heated for 15 min and left to cool to RT for 15 min. Slides were then washed with distilled water and endogenous peroxidase was blocked for 30 min with 3% H_2O_2 (diluted in methanol). Sections were washed with PBST (PBS + 0.1% Tween-20) and blocked for 1 h in 10% goat serum in PBS. Slides were incubated for 1 h at RT with rabbit polyclonal anti-CTCF antibody (1:500) (Soshnikova et al. 2010). Then, slides were washed with PBST twice for 5 min and incubated for 30 min at RT with a biotinylated secondary antibody goat anti-rabbit (1:300 diluted in PBST). Slides were washed with PBST twice for 5 min and incubated for 30 min at RT with Vectastain Elite ABC Kit (reference vector VC-PK-6100-KI01), washed three times for 5 min with PBST and developed with the Peroxidase Substrate Kit for DAB (reference Vector SK-4100). Slides were washed with distilled water, dehydrated through an increasing ethanol series (70%, 95%, 100% 2 min each), cleaned with xylene (twice for 5 min) and mounted in Dpx Mountant medium (Sigma 06522).

2.7. Phosphohistone 3 (PH3) and TUNEL staining

PH3 and TUNEL staining was performed in 5-μm-thick sections. The protocol is designed to perform both staining on the same samples at once. The protocol starts with the TUNEL, then is followed by PH3 staining and ends up with the incubation of secondary antibodies for each staining at the same time. Slides were incubated overnight at 37ºC and then incubated at 65ºC and dewaxed with xylene twice for 5 min. They were hydrated three times with 100% ethanol, once with 95% ethanol, and once with water, 5 min each. Then, they were rinsed with distilled water three times and antigens were retrieved with 10 mM citric acid pH 6.0 as before. Slides were then washed with distilled water and endogenous peroxidase was blocked for 40 min with 1% H_2O_2 (diluted in methanol) in the dark. They were then rinsed quickly with distilled water three times and washed once with distilled water for 5 min. Incubation with 0.5% Triton X-100 diluted in PBS for 10 min at RT was followed by a PBS rinse three times and then a PBS wash for 5 min. From this step onwards, slides were incubated in a humidified chamber in the dark. They were incubated with 1mM $CoCl_2$ in TdT buffer for 15 min at RT followed by incubation with TdT mix (appendix A) for 1 hour at 37ºC. Then, slides were washed with 1mM $CoCl_2$ in TdT buffer for 10 min. The reaction was stopped by washing in 0.01% Tween 20 diluted in PBS twice for 5 min. This was followed by three 5 min washes in PBS. Next part is the PH3 staining. Slides were

blocked with Histoblock for 1 hour at RT, followed by incubation with an anti-PH3 antibody (1:200, Millipore, reference 06-570) for 1 hour at 37ºC. They were washed three times with PBS for 5 min. This is the final part for both stainings. Incubation with secondary antibody Alexa488 goat anti-rabbit (1:500, Molecular Probes, reference A11034) and streptavidin-Cy3 (1:500 Jackson, reference 016-160-084) for one hour at RT, was followed by three washes with PBS for 5 min. Slides were incubated for 10 min at RT with 4',6-diamidino-2-phenylindole (Dapi, Millipore 1246530100), washed three times with PBS for 5 min and finally mounted in Dpx Mountant medium.

Quantification of TUNEL and PH3 staining were made with ImageJ software. Positive cells for each antibody were counted in 3–5 sections per heart. Three wild-type and three KO hearts were used. Statiscal significance was determined using one-tailed Student´s test from GraphPad.

2.8. Imaging

Hematoxylin and eosin staining, RNA *in situ* hybridization in sections, TUNEL and PH3-immunostained slides were observed with an Olympus BX51 microscope and photographed with an Olympus DP71 digital camera and CellSens Entry Software. Whole-mount embryos processed for whole-mount RNA *in situ* hybridization and β-galactosidase staining were observed using a Leica MZ FLIII Scope and photographed with a Nikon digital camera DMX 1200F and NIS-Elements D 3.2 software. Whole-mount embryos fixed to photograph the *Ctcf* KO phenotype in Fig. 11 and Fig. 12 were observed with a Leica MZ FLIII Scope and photographed with an Olympus DP71 digital camera and NIS-Elements D 3.2 software.

3. RNA-seq and data analysis

3.1. Sequencing and analysis

RNA-seq was performed on three pools of six E10.5 hearts each from control (*Ctcf*$^{fl/+}$), heterozygotes (*Ctcf*$^{fl/+}$*; Nkx2.5-cre*) and homozygous mutants (*Ctcf*$^{fl/fl}$*; Nkx2.5-cre*). Individual hearts were frozen in liquid nitrogen and stored at -80ºC. The yolk sac of each embryo was used for posterior genotyping. RNA was extracted with the miRNeasy Mini Kit (Qiagen, reference 217004).

Sequencing was performed at the CNIC Genomics Unit using the GAIIx sequencer using a 75 bp single end elongation protocol. Reads were aligned with Bowtie aligner (v0.12.7) using the transcriptome set from Mouse Genome Reference NCBIM37 and Ensembl Gene Build version 65. Then, they were quantified RSEM v1.20. Genes with

less than one count per million in more than three out of nine samples were filtered out. Thus, 14011 genes were considered for the differential expression analysis. Data were then normalized using a correction factor implemented in the bioconductor package EdgeR and described by Robinson & Oshlack (Robinson and Oshlack 2010). Finally, genes with an adjusted p-value <=0.05 were considered as differentially expressed.

The matrix of normalized expression values was used to compute a Euclidian distance between samples and represented using heat map function. The heatmap (Fig. 8) shows that the control and the heterozygotes are in close proximity, indicating that they are similar.

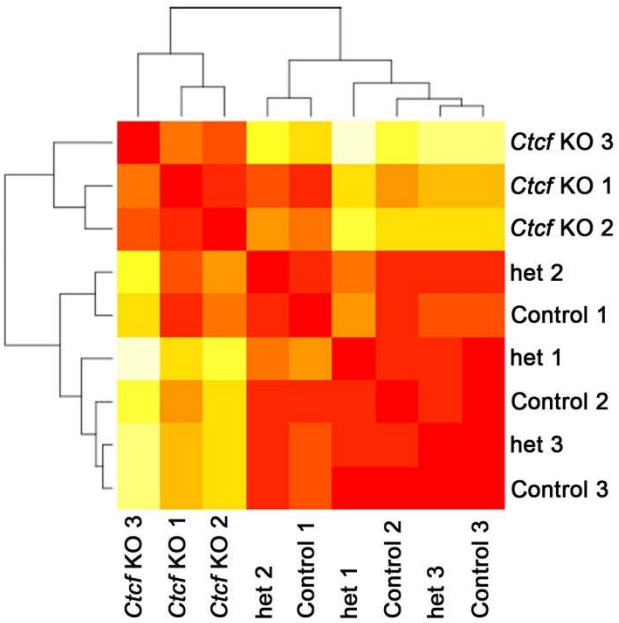

Fig. 8. Heat map with the samples and replicates used in the RNAseq.

This representation shows that the control and the heterozygotes are more similar to each other and more different from the *Ctcf* KO. het = heterozygote.

3.2. CTCF binding sites and heart enhancer analysis

Distances between transcriptional start sites (TSS) of the DEGs (from the CTCF KO versus control comparison) and CTCF binding sites or predicted heart enhancers were analyzed using R statistical software. Coordinates for 8-week-old and E14.5 hearts CTCF binding sites and predicted heart enhancers were obtained from the Ren Laboratory at UCSD (http://chromosome.sdsc.edu/mouse/download.html) (Shen et al. 2012)). For the analysis to determine if the TSS of the DEG genes were closer to a CTCF binding site or a predicted heart enhancer we used the Mann-Whitney test, that it is used for non-parametrics and independent datasets. For the analysis to determine

if the TSS of the DEG were closer to a CTCF binding site or a predicted heart enhancer in the 10 and 20 Kb windows, we used the Proportion test.

3.3. Gene density analysis

Gene density analysis was done using R statistical software. Distances between the transcriptional start sites (TSS) of the DEGs (from the *Ctcf* KO versus control comparison) and the first neighboring gene to both sides was measured in Kb. These two distances were added, and the sum of that was considered the total distance for each gene. Then, the median of the total distance was compared between upregulated and downregulated genes. To analyze if the difference between medians was significant, we used the Wilcoxon test, for non-parametric dependent datasets. The Density curves show the distribution of downregulated and upregulated genes in mutant hearts in relation to the distance from the TSS of adjacent 5' and 3' genes.

3.4. Heat maps

Heat maps to represent RNAseq differential expression several genes in Fig. 19 and 28 were generated using heatmap.2 package function from G plots from the R package.

3.5. Gene ontology

Gene ontology (GO) enrichment analysis was performed using DAVID (http://david.abcc.ncifcrf.gov) with cut-offs of 10 genes and EASE value of 0.001. The EASE value is a modified Fisher Exact P-Value, for gene-enrichment analysis. The GO terms presented in this work are those included in the group 5 of biological process. In table 5 are the details for each column. The GO terms are ordered from high- to low-fold enrichment. Further details from the GO term tables are shown in Table 5.

Table 5. Gene Ontology columns description.

Term	Count	%	P-value	Fold enrichment
Enriched GO term associated with our gene list.	Genes involved in the term. We established that each GO term should have at least 10 genes.	Percentage of involved genes in the GO term. 100% is the total genes from the list.	Modified Fisher Exact P-value, EASE Score. The smaller the score, the more enriched.	Defined as the ratio of two proportions. The proportion of genes from our list enriched in a particular GO term, divided by the proportion of genes from the background enriched in the same particular GO term.

4. Mouse transgenic assays

4.1 Cloning

Mouse DNA fragments 1-7 (#1 chr13:73,307,200-73,308,500; #2 chr13:73,312,500-73,318,000; #3 chr13:73,400,238-73,401,992; #4 chr13:73,414,400-73,416,300; #5 chr13:73,418,664-73,422,000; #6 chr13:73,427,100-73,428,900; #7 chr13:73,435,722-73,438,701; Mouse Genome Reference NCBIM37, mm9) were amplified by PCR (primers used are detailed in Table 6 and PCR conditions are detailed in Table 7) from genomic DNA of CD1 adult mice. PCR products were cloned in the pGEMTeasy vector (Promega, reference A1360). Fragments were then isolated from pGEMTeasy by NotI digestion and subcloned into the *Not*I site of the p1230 vector (Fig. 9). The tested DNA fragments together with the β-globin-LacZ

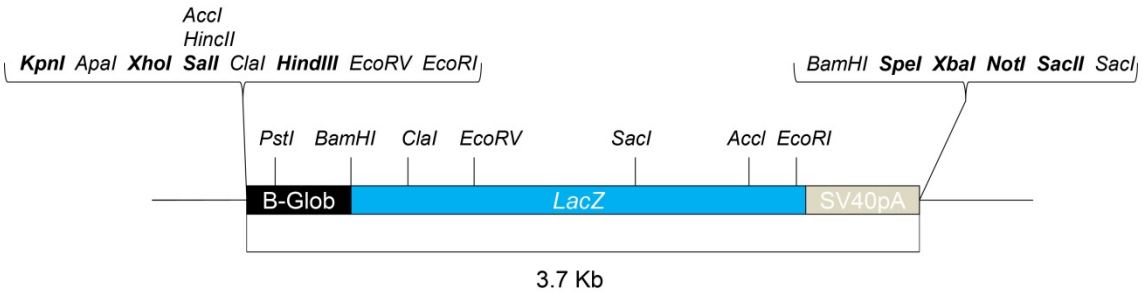

Fig. 9. Map of the p1230 vector.

The p1230 vector contains a human β-globin minimal promoter, the *E. coli* β-galactosidase (*LacZ*) gene and a SV40 polyadenylation signal, depicted in black, blue and grey, respectively. Unique cloning sites are shown in bold.

cassette and SV40pA were released from the vector with *SalI* and *SacII*. Only fragment #2 was released with *SalI* and *BgmBI*, because this fragment had a *SacII* restriction site in its sequence (*BgmBI* is not present in the p1230 multicloning site). After release, fragment 2 looses 11 bp. Digested DNA was run in 1% low gelling temperature agarose (Sigma, reference A9414), and the fragment of interest was extracted with the QIAquick Gel Extraction Kit (Qiagen, reference 28704) and re-suspended in oocyte water (Sigma-Aldrich, reference W1503-100ML).

Table 6. Primers for amplification of genomic fragments for transgenic assays

Primer	Sequence
frag1_F	CCTGCATTGCTCTGCTGTTTAA
frag1_R	TGGTCAGAAGGCAAATCTGAAGG
frag2_F	CCAGTCAGGCAATGTCTGTGTC
frag2_R	ACCATGTTAGGAGAGTGTTTCC
frag3_F	GCCAGACGGAGCAGTGGCATG
frag3_R	TGGTCAGAAGGCAAATCTGAAGG

frag4_F	GGATAGCCAAACGGCCAGGC
frag4_R	GTGCTCTTAACTGCTGAGCCAT
frag5_F	GGTCCCACTCTAAAGCCCAGAA
frag5_R	GGAGGTCATGAAGAGATTGAATC
frag6_F	GACCAGCCTGGAACTCACAAC
frag6_R	GATGACTGAGTGGCACTCTGC
frag7_F	CCATTTCTAGAGAAGGCTGGTT
frag7_R	ACCAAGGTTGTCTCTGTGGG

F stands for forward primer and R stands for reverse primer

Table 7. PCR conditions for amplification of genomic fragments for transgenic assays

Initial Denaturation	95ºC	5 min	
Denaturation	94ºC	30 sec	35 cycles
Annealing	59ºC	30 sec	
Elongation	72ºC	2 min	
Final Elongation	72ºC	10 min	
Hold	4ºC	∞	

4.2. Microinjection and embryo transfer

Transgenic mouse embryos were generated as previously described (Aguirre et al. 2015; Nagy et al. 2003). DNA was injected into the pronucleus of E0.5 B6/CBA mice at 4–6 ng/µL diluted in microinjection buffer (appendix A). Embryos were then transferred to CD1 pseudo-pregnant females. Embryos were collected from E9.5 to E12.5, fixed and stained for β-galactosidase activity. Negative embryos for β-galactosidase activity were genotyped for *LacZ* and myogenin as an internal control by PCR to calculate transgenic efficiency. Primer details for *LacZ* and myogenin are shown in Table 8 and PCR conditions are shown in Table 9.

Table 8. Primers for Transgenic Mouse Embryos

Primer	Sequence	Notes
lacZ_F	GCGACTTCCAGTTCAACATC	genotyping of transgenic embryos
lacZ_R	GATGAGTTTGGACAAACCAC	genotyping of transgenic embryos
Myogenin_F	CCAAGTTGGTGTCAAAAGCC	control for transgenic genotyping
Myogenin_R	CTCTCTGCTTTAAGGAGTCAG	control for transgenic genotyping

F stands for forward primer and R stands for reverse primer

4.3. β-galactosidase staining

Embryos were fixed with LacZ fix (37% formaldehyde, 2 mM $MgCl_2$, 5 mM EGTA (Ethylene-bis(oxyethylenenitrilo)tetraacetic acid, glycol ether diamine tetraacetic acid,

reference Acids E3889), 0.2% glutaraldehyde) for 20 min at RT, washed with PBS + 0.2% Tween for 20 min at RT and stained overnight with LacZ stain: 5 mM K_3Fe, 5 mM K_4Fe, 2 mM $MgCl_2$, 1.04 mg/mL X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, Sigma reference B4252).

Table 9. PCR conditions for amplification of *LacZ* and *myogenin*.

Initial Denaturation	95ºC	5 min	
Denaturation	94ºC	1 min	30 cycles
Annealing	60ºC	1 min	
Elongation	72ºC	1 min	
Final Elongation	72ºC	10 min	
Hold	4ºC	∞	

5. Circularized Chromosome Conformation Capture: 4C

5.1. Tissue collection and fixation

Pools of 45-65 hearts from controls or mutants were lysed and 4C was performed as previously described (Splinter et al. 2012; van de Werken et al. 2012a). Hearts were dissected from control (*Ctcf*$^{fl/+}$ or *Ctcf*$^{fl/fl}$) and mutant (*Ctcf*$^{fl/fl}$*; Nkx2.5-cre*) E11.5 embryos. This stage was used in order to obtain enough starting material. Each heart was incubated with 48 U/µL of collagenase type II (Worthington, reference 4176) at 37ºC for 45 min, minced with a p200 pipette tip and passed through a 70 µm cell strainer. Then, samples were centrifuged for 1 min at 2600 rpm and the supernatant was discarded. Samples were re-suspended with 600 µL of PBS and an equal volume of 4% PFA was added. Samples were then crosslinked for 10 min at RT with 2% PFA while agitating. Then 171 µL of cold glycine (final concentration 0.285 mM) was added and mixed by inversion. Samples were centrifuged for 8 min at 1300 rpm and 4ºC, the supernatant was discarded and samples were frozen with liquid nitrogen and stored at -80º for later processing once genotypes were established.

5.2. Cell lysis

Pools of hearts were re-suspended in 5 mL of cold 4C buffer with douncing on ice to increase lysis efficiency. Three microliters of disaggregated hearts were mixed with 3 µL of Methyl Green-Pyronin stain (Sigma, reference HT70116) on a slide, which was overlayed with a coverslip to check cell lysis. Under the microscope, cytoplasm stains pink and nuclei stains blue/green. Once cell lysis was completed, samples were centrifuged for 5 min at 1800 rpm at 4ºC and supernatants were discarded.

5.3. First digestion

Samples were re-suspended in 450 µL of MiliQ water with the addition of 60 µL of 10× *DpnII* restriction enzyme buffer (New England Biolabs, reference R0543M). Tubes were placed at 37ºC and 15 µL of 10% SDS was added followed by incubation at 37ºC and 900 rpm in the Ependdorf Thermomixer for 1 h. Next, 75 µL of Triton X-100 was added and incubation was continued as before for 1 h. From here, 5 µL of sample was removed as an undigested control. Then, 400 U of *DpnII* was added and samples were incubated as before overnight. On the next day, 200 U of *DpnII* was added and samples were incubated for a further 4 h. From here, 5 µL was removed as the first digestion control. To determine digestion efficiency, 92.5 µL of 10 mM Tris-HCl pH7.5 was added to the 5 µL undigested and first digestion controls followed by addition of 2.5 µL of proteinase K (20 mg/mL), and incubation for 1 h at 65ºC. Then, 1 volume (100 µL) of phenol-chloroform was added, tubes were vigorously shaken and centrifuged at 13000 rpm for 5 min at RT. The aqueous phase was recovered and 20 µL was loaded onto a 0.6% agarose gel to visualize different sizes of DNA, “a smear” generated after digestion.

5.4. First Ligation

When the first digestion was completed, *DpnII* was heat inactivated at 65ºC for 20 min. A ligation reaction was set up in a final volume of 1500 µL (150 µL of 10× ligase buffer, 60 U of T4 DNA ligase (Promega, reference 10481220001), ~500 µL of first digestion and water to 1500 µL) and incubated overnight at 4ºC. The next day, 50-100 µL was taken as first ligation control, to which 2.5 µL of proteinase K (20 mg/mL) was added followed by incubation for 1 h at 65ºC. Then, 1 volume of phenol-chloroform (50 or 100 µL) was added, tubes were vigorously shaken and centrifuged at 16400 g for 10 min at RT. The control samples were then loaded onto a 0.6% agarose gel to visualize the shift in DNA size “smear” as a result of the ligation.

5.5. Reverse crosslinking and DNA purification

When the first ligation was completed, 15 µL of proteinase K (20 mg/mL) was added and samples were incubated overnight at 65ºC to reverse crosslinking. The next day 30 µL of RNAse (10 mg/mL) was added and samples were incubated for 45 min at 37ºC. The 1500 µL sample was then transferred to a 15 mL falcon tube and 1 volume of phenol-chloroform was added and samples were vigorously shaken. Following centrifugation at 3270g for 15 min at RT, the aqueous phase was transferred to a new 15 mL falcon tube. Then, 1/10 volume of 3 M NaAC pH 5.6 (150 µL), 2.5 volumes of 100% ethanol (3750 µL) and 20 µL of glycogen (Roche, reference 10901393001) were added, mixed by inversion and the mix was distributed into 1500 µL Eppendorf tubes,

which were incubated at -80ºC until samples were frozen (2-3 h). Samples were centrifuged at 10,000 rpm for 20 min at 4ºC, pellets were washed with cold 70% ethanol, centrifuged again as before, air dried, and dissolved in 150 µL of 10 mM Tris-HCl pH 7.5 ~15 min at 37ºC and left overnight at 4ºC.

5.6. Second Digestion

To the 150 µL of the first digestion samples, 50 µL of 10× *Csp6I* buffer (Fermentas, reference ER0211), 50 U of *Csp6I* and MiliQ water up to 500 µL were added. Samples were incubated overnight at 37ºC and then 5 µL was removed as the second digestion control. Then, 95 µL of 10 mM Tris-HCl pH7.5 was added and 20 µL of all controls were loaded in a 0.6% agarose gel.

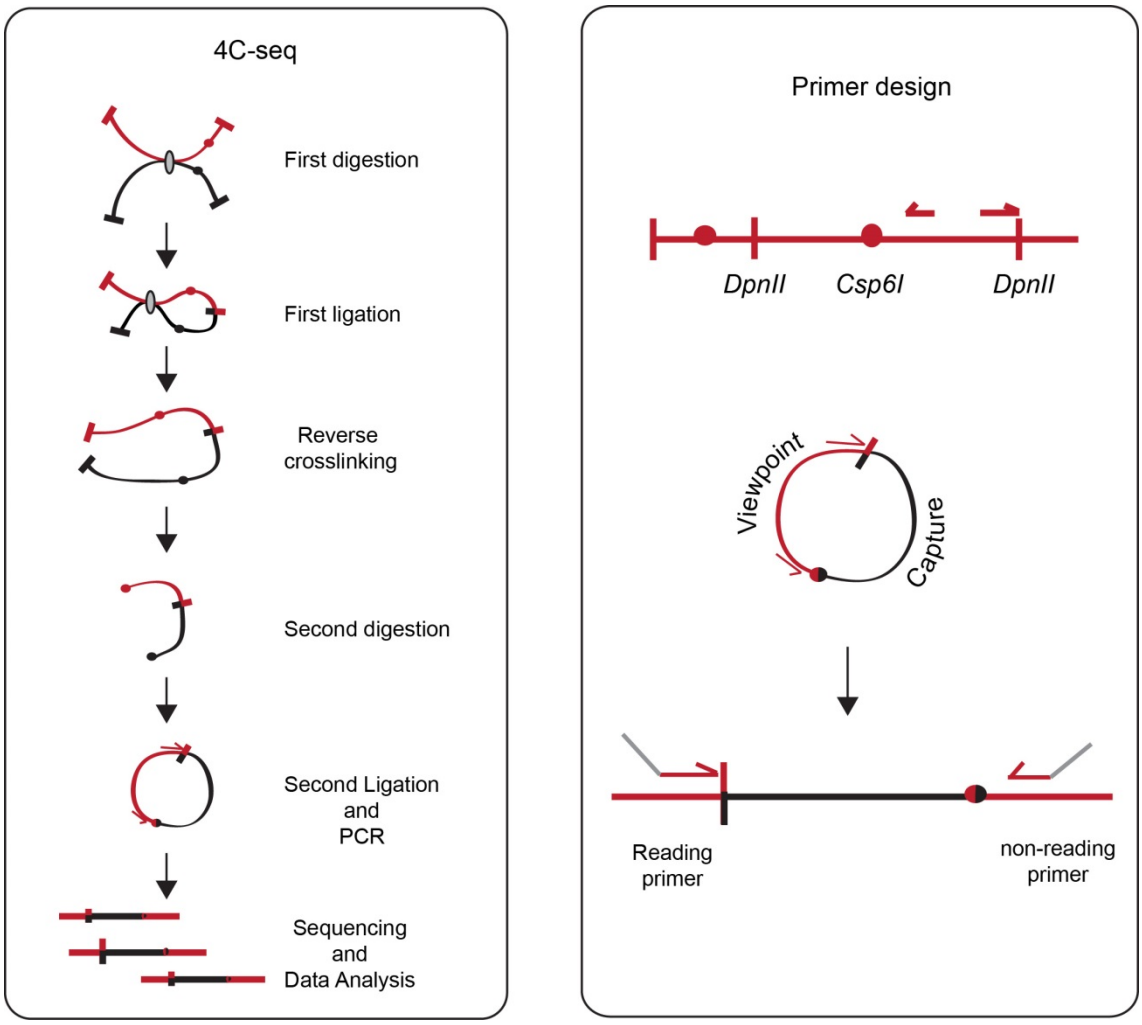

Fig. 10. 4C seq.

The left panel shows the graphical summary of the 4Cseq procedure. The right panel shows the logic for primer design: on the top, the primers face outward from the viewpoint and are depicted in red. The orientation is outward to perform inverse PCR. In the bottom of the right panel, the gray overhangs in the reading and non-reading primers represents the Illumina adaptors. Scheme was modified from (Splinter et al. 2012).

5.7. Second ligation and DNA purification

When the second digestion was completed, *Csp6I* was heat-inactivated at 65ºC for 25 min. A ligation reaction was set up in a final volume of 1500 µL (150 µL of 10× ligase buffer, 100 U of T4 DNA ligase, ~500 µL of second digestion and water to 1500 µL) and incubated overnight at 4ºC. The next day, 50-100 µL was removed as the second ligation control, and 20 µL of all controls were loaded in a 0.6% agarose gel; in total, 5 control samples were loaded in a 0.6% agarose gel to visualize the shift in DNA spread. DNA was purified as for the second digestion. The pellet was re-suspended in 50 µL of 10 mM Tris-HCl pH 7.5. Samples were purified with the QIAquick PCR purification kit (Qiagen, reference 28104).

5.8. Primers and viewpoints

Irx1, *Irx2* and *Irx4* primers were designed in the laboratory of Jose Luis Gómez Skarmeta, (CABD at Seville, Spain). Other primers were designed using the mouse mm9 version of the UCSC Genome Browser. The viewpoint is the DNA sequence between two restriction sites from the first enzyme, in our case *Dpn*II. The ideal size is 500 bp. The DNA sequence between the first and the second (*Csp*6I in our case) restriction enzyme should be at least 300 bp to allow circularization (van de Werken et al. 2012a). The reading primer is designed on top of the first restriction enzyme site with a 20 nt length. The non-reading primer design is more flexible and can be 100 bp from the second restriction enzyme site and with a length between 18 and 25 nt. Once designed, short primers were tested with decreasing DNA concentrations (100, 50, 25 and 12.5 ng). Once checked, Illumina adaptors were added to the 5´end of the short primers so the PCR product is ready to sequence. P5 and P7 adapters are used for the reading and non-reading primer respectively (Splinter et al. 2012; van de Werken et al. 2012a).

5.9. PCR conditions

For all experiments, 100 to 200 ng of the resulting 4C template was used for PCR with the Expand Long Template Polymerase (Roche, reference 11759060001) (primers used are detailed in Table 10 and PCR conditions are in Table 11). A graphical summary of the 4C-seq is in Fig. 10.

Table 10. 4C viewpoint primers

Primer	Sequence
*Irx1*_DpnII	TCTCTGCTCTGGGTTGGATC
*Irx1*_Csp6	TTGGGGCTCTCTGTCAAAC
*Irx2*_DpnII	GATGAGGGGTGCCCGATC
*Irx2*_Csp6	CTGGGCATCCCACTTCTACA

*Irx4*_DpnII	TCCGGCGCAAGAGCGATC
*Irx4*_Csp6	GTAAGCGGATGGGAAGGAC
*Ndufs6*_DpnII	AGCATGTGTTGTTTGGGATC
*Ndufs6*_Csp6	AACACTGCGCACAGTTAAGC
CTCF BS_DpnII	TGGAACCCACAGTGCTGATC
CTCF BS_Csp6	ACCAGCAAATTAATCTACAAGGC
Frag1_DpnII	ATGTAATGTTGCTGGAGATC
Frag1_Csp6	GCACTTAGGATGTGTTGACCC
Adaptor reading primer	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT
Adaptor non reading primer	CAAGCAGAAGACGGCATACGA

The reading primers have *Dpn*II and non-reading primers have *Csp*61 restriction sites

Table 11. PCR conditions for 4C-seq library amplification

Initial Denaturation	94ºC	2 min	
Denaturation	94ºC	10 sec	30 cycles
Annealing	55ºC	1 min	
Elongation	68ºC	3 min	
Final Elongation	68ºC	5 min	
Hold	1º2ºC	∞	

5.10. Sequencing and data analysis

Sequencing was performed at the CNIC Genomics Unit using the GAIIx Illumina and the HiSeq 2000 Sequencer. Then raw sequencing data was de-multiplexed using the reading primer as barcode to identify the reads (captures) of each viewpoint. Reads were aligned to an *in silico* library generated based on the restriction enzymes used (Noordermeer et al. 2011; Splinter et al. 2012; van de Werken et al. 2012a), in our case *DpnII* and *Csp6I*. The mouse mm9 reference genome was used for the *in silico* library. For sample comparison between controls and mutants, the 4C data were normalized by total weight in a window of 15 Mb surrounding the *IrxA* cluster. The mean and standard deviation for each fragment were calculated for each group of replicates and the difference was determined between the means in control and mutants for each viewpoint. To assess whether these differences were statistically significant, the entire region was divided into sliding windows of 50 *Dpn*II-*Csp*6I restriction fragments with an 80% overlap and the number of reads in each window for each sample was counted. Data were compared using the R package edgeR (Robinson et al. 2010), which applies statistical tests based on negative binomial distributions.

6. Genome editing with the CRISPR/Cas9 system

6.1. Guide RNAs

Guide RNAs were designed with an online tool (http://crispr.mit.edu/). Details of the oligonucleotides are shown in Table 12. The T7 promoter sequence was added at the 5´ end of each oligo and they were transcribed with the mMESSAGE mMACHINE® T7 Ultra Kit Synthesis of Translation Enhanced Capped Transcripts (Thermofisher, reference AM1345).

Table 12. Oligos for guide RNAs

Oligos	Sequence	Location
g5.3	GGCGTCCAATTGACAAATTG / PAM: TGG	chr13:73297211-73297230
g5.4	AATTTGAGTTCTCTCCGAAC / PAM: AGG	chr13:73297805-73297825
g3	CTTGTCCGTGCGGTCCAAC / PAM: TGG	chr13:73298566-73298585

6.2. Microinjection, embryo transfer, collection and processing

Two different combinations of guide RNAs (g5.4-g3 and g5.3-g3) together with a Cas9 protein with a nuclear localization signal (PNA Bio reference CP01) were injected into the pronucleus of E0.5 B6/CBA F1 embryos. Guide RNAs were injected at 20–30 ng µL, and the Cas9 protein was injected at 6–72 ng/µL. Embryos were then transferred to CD1 pseudo-pregnant females, collected at E9.5-10.5 and processed individually for *in situ* hybridization and genotyping. Primers for genotyping and PCR conditions are described in Table 13 and 14, respectively. The PCR strategy was designed to distinguish the unaltered and the deleted allele by PCR product size. The unaltered allele was 1608 bp. The g5.3 and g3 combination produced a 1374 bp deletion and the PCR product size of the deleted allele using this guide RNA combination was 234 bp. The g5.4 and g3 combination produced a 779 bp deletion and the PCR product size of the deleted allele using this guide RNA combination was 829 bp. The size of the PCR products of two deletions, can vary since the deletion, although highly consistent was not exactly the same in each embryo. The PCR products of deleted alleles were cloned into a pGEMTeasy vector, and further sequenced to confirm that the desired deletion took place.

Table 13. Primers for deletion using the CRISPR/Cas9 system

Primer	Sequence
CTCFdel_F	AGGCCTTTAGCAGGAACCTC
CTCFdel_R	CGATCCTGGAACAGGAAAAA

F stands for forward primer and R stands for reverse primer

Table 14. PCR conditions for amplification of the CTCF binding site deletion.

Initial Denaturation	95ºC	5 min	
Denaturation	94ºC	1 min	35 cycles
Annealing	58,6ºC	1 min	
Elongation	72ºC	1min,50sec	
Final Elongation	72ºC	10min	
Hold	4ºC	∞	

RESULTS

1. Analysis of the *Ctcf* KO mouse embryonic heart phenotype

1.1. *Ctcf* deletion in the heart results in embryonic lethality

To investigate the role of CTCF during heart development, we used *Cre* recombinase driven by the cardiac homeobox gene promoter *Nkx2.5* (*Nkx2*.5-*Cre*) to delete *Ctcf* in the embryonic mouse heart. This driver is functional as early as E7.5, when cardiac specification takes place (Stanley et al. 2002), and is expressed in the myocardium and endocardium leading to deletion in both myocardial and endocardial cells. From the crosses established to delete *Ctcf* in the heart, four possible genotypes are predicted. Animals homozygous for the *Ctcf* heart deletion (*Ctcf*$^{fl/fl}$*;Nkx2.5-Cre,* referred to in this work as *Ctcf* KO), heterozygous animals, which had only one *Ctcf* allele deleted (*Ctcf*$^{fl/+}$*;Nkx2.5-Cre*), and controls that were either homozygous or heterozygous for the *Ctcf* floxed allele (*Ctcf*$^{fl/fl}$ or *Ctcf*$^{fl/+,}$ respectively). Controls did not carry the *Nkx2.5-Cre* allele (Fig.11A appendix B). In theory, we expected to obtain 25% of each genotype. After weaning, we genotyped the litters and we failed to recover any *Ctcf* heart-specific KO (*Ctcf*$^{fl/fl}$*;Nkx2.5-Cre*), what was indicative of embryonic or perinatal lethality. Also, that one *Ctcf* allele is enough for proper heart development. To establish if embryo death was occurring, we collected embryos at different developmental stages. We recovered live *Ctcf* KO embryos up to E12.5 (Fig. 11A), whereas E13.5 embryos were dead (Fig. 11B,C). Furthermore, E13.5 embryos were not fully developed and were smaller in size than controls (Fig. 11B,C). We therefore conclude that deletion of *Ctcf* in the heart leads to embryonic lethality at E13.5.

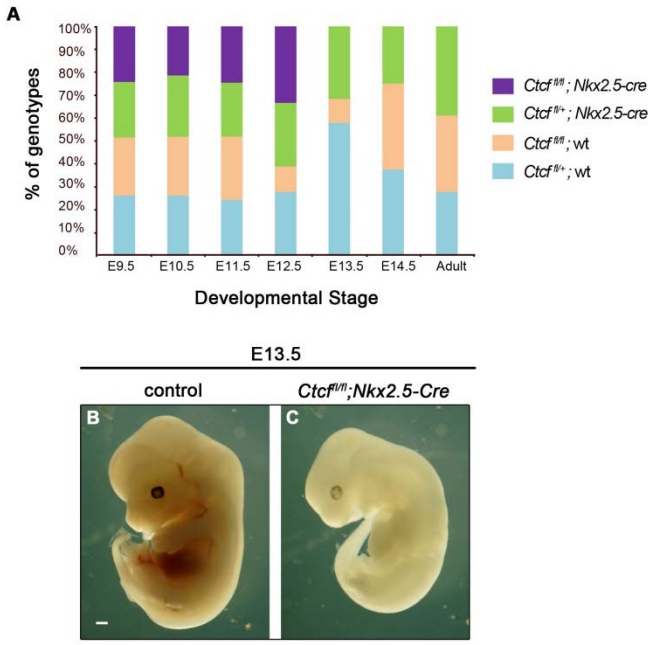

Fig. 11. *Ctcf* deletion in the heart is embryonic lethal.

A, Graph showing the percentage of embryos obtained of each genotype. **B, C,** At 13.5, the heart *Ctcf* KO embryo is dead, delayed in development and smaller than the control. Scale bar is 500 µm.

1.2. *Ctcf* deletion in the heart leads to cardiac malformations

We next examined the general morphology of the *Ctcf* KO embryos and hearts at stages prior to death. The external appearance of the whole embryo at E11.5 was similar to that of the control (Fig. 12A,B), and this was also the case at earlier stages. When we examined E12.5 embryos, morphological differences were detected in the whole embryo, including pericardial edema (Fig. 12E-H).

We removed the pericardium of these embryos to visualize the heart. The E11.5 *Ctcf* KO heart presented a reduction in size (Fig. 12C,D), which was exacerbated at E12.5 (Fig. 12I,J). Embryos at these stages were alive, but presented heart malformations that very likely contributed to the lethality.

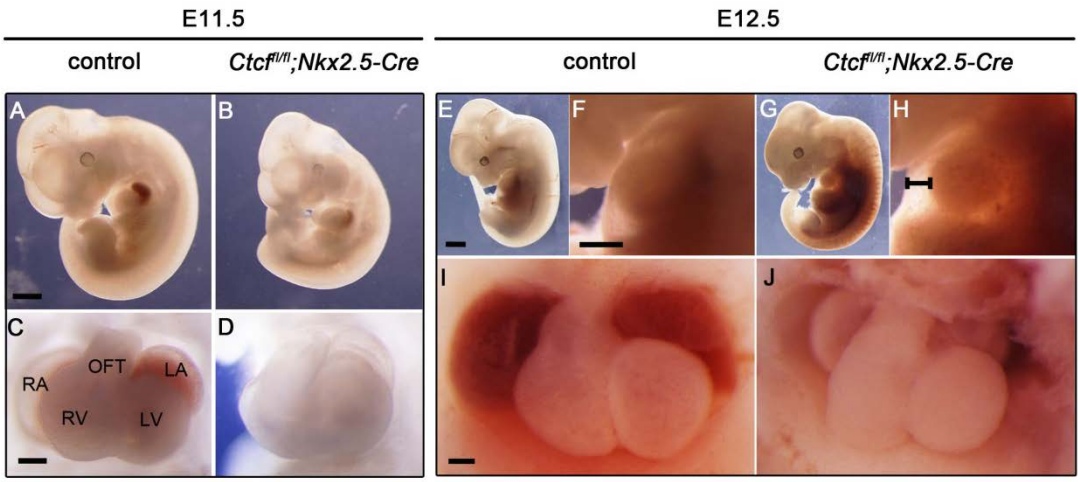

Fig. 12. *Ctcf* deletion in the heart alters heart development.

A, Control embryo at E11.5. **B,** *Ctcf* KO embryo has no obvious morphological defects. **C,D,** The *Ctcf* KO E11.5 heart is reduced in size. **E,** Control embryo at E12.5. **F,** E12.5 thorax of control embryo. **G,** *Ctcf* KO embryo at E12.5. **H,** The *Ctcf* KO thorax shows a reduction in the size of the heart. **I,J,** The *Ctcf* KO E12.5 heart is smaller than the control. Scale bar in A, E and F is 500 µm. Sale bar in C and I is 200 µm. RA: right atrium, RV: right ventricle, OFT: outflow tract, LA: left atrium, LV: left ventricle.

To examine the morphology in more detail, we performed hematoxylin and eosin (H&E) staining and CTCF immunostaining on E11.5 and E12.5 embryos. When KO embryos were compared with the control embryos, H&E staining of *Ctcf* KO embryos at E11.5 revealed gross cardiac malformations that included a shorter and broader interventricular septum (IVS) (Fig. 13A,B) and a thinner compact myocardium (Fig. 13C,D). CTCF immunostaining at E11.5 (Fig. 13E,F) showed that CTCF was absent in the *Nkx2*.5 territories (Fig. 13G,H arrows) but was present in territories where *Nkx2*.5 was not expressed (Fig. 13G,H arrowheads pointing to the body wall) (Stanley et al. 2002), confirming cardiac deletion of *Ctcf* in KO embryos. At E12.5, H&E staining also highlighted exacerbated gross cardiac malformations and the reduction in heart size was conspicuous (Fig. 14A-D). CTCF immunostaining at E12.5 demonstrated downregulation of the protein (Fig. 14E-H).

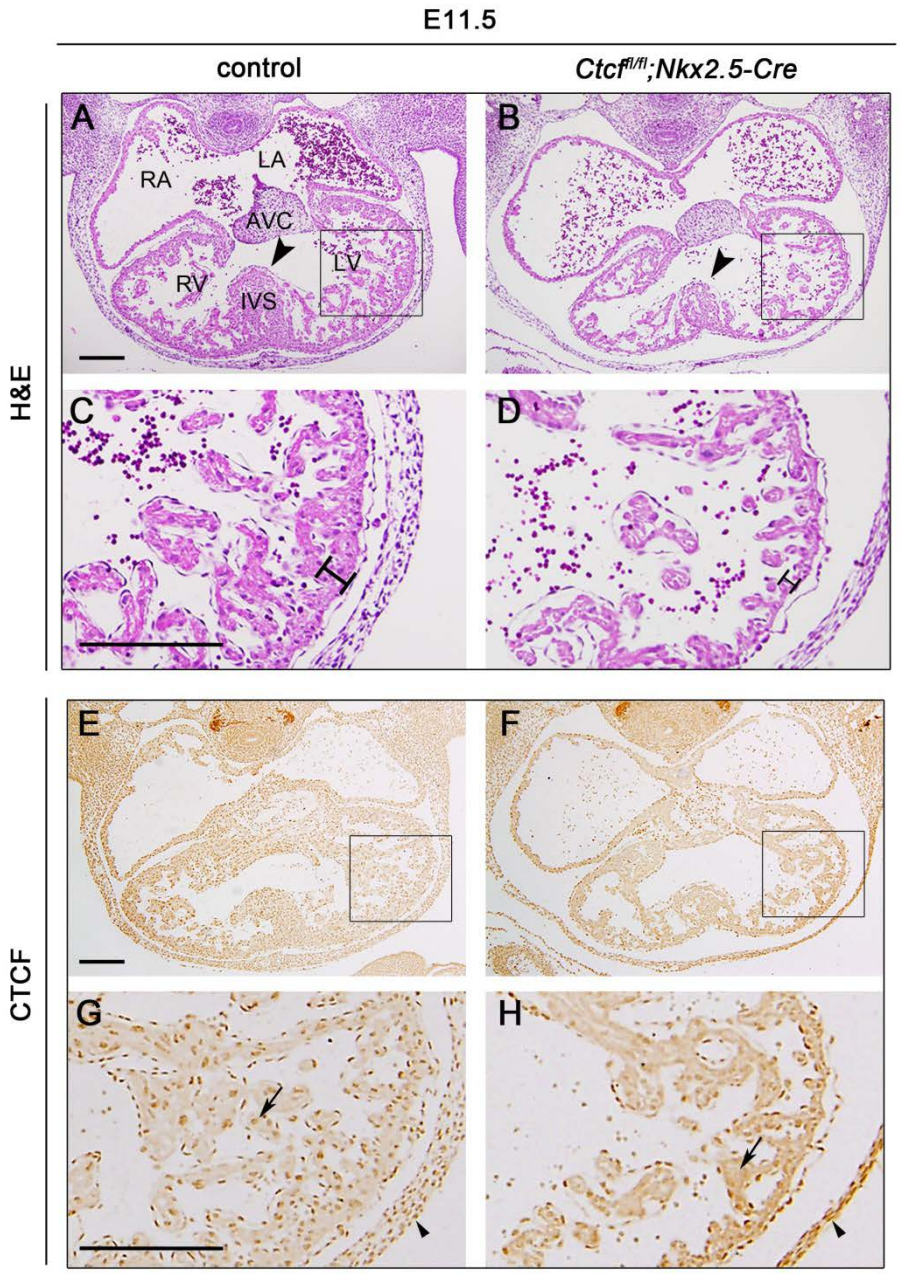

Fig. 13. *Ctcf* deletion in the heart causes gross cardiac malformations at E11.5.

A-D, H&E staining in transverse sections of control and *Ctcf* KO E11.5 hearts. *Ctcf* KO shows a deformed IVS (arrowheads in **A** and **B**) and a thinner myocardium (bracket in **C** and **D**). **E-H,** CTCF immunostaining shows reduced CTCF expression in the KO heart (**E**,**F**). This downregulation is better appreciated in **G** and **H**. Arrows show *Nkx2.5* territories and arrowheads show non-*Nkx2.5* territories, the pericardium. RA: right atrium, RV: right ventricle, IVS: interventricular septum, AVC: atrioventricular canal, LA: left atrium, LV: left ventricle. Scale bars are 200 µm.

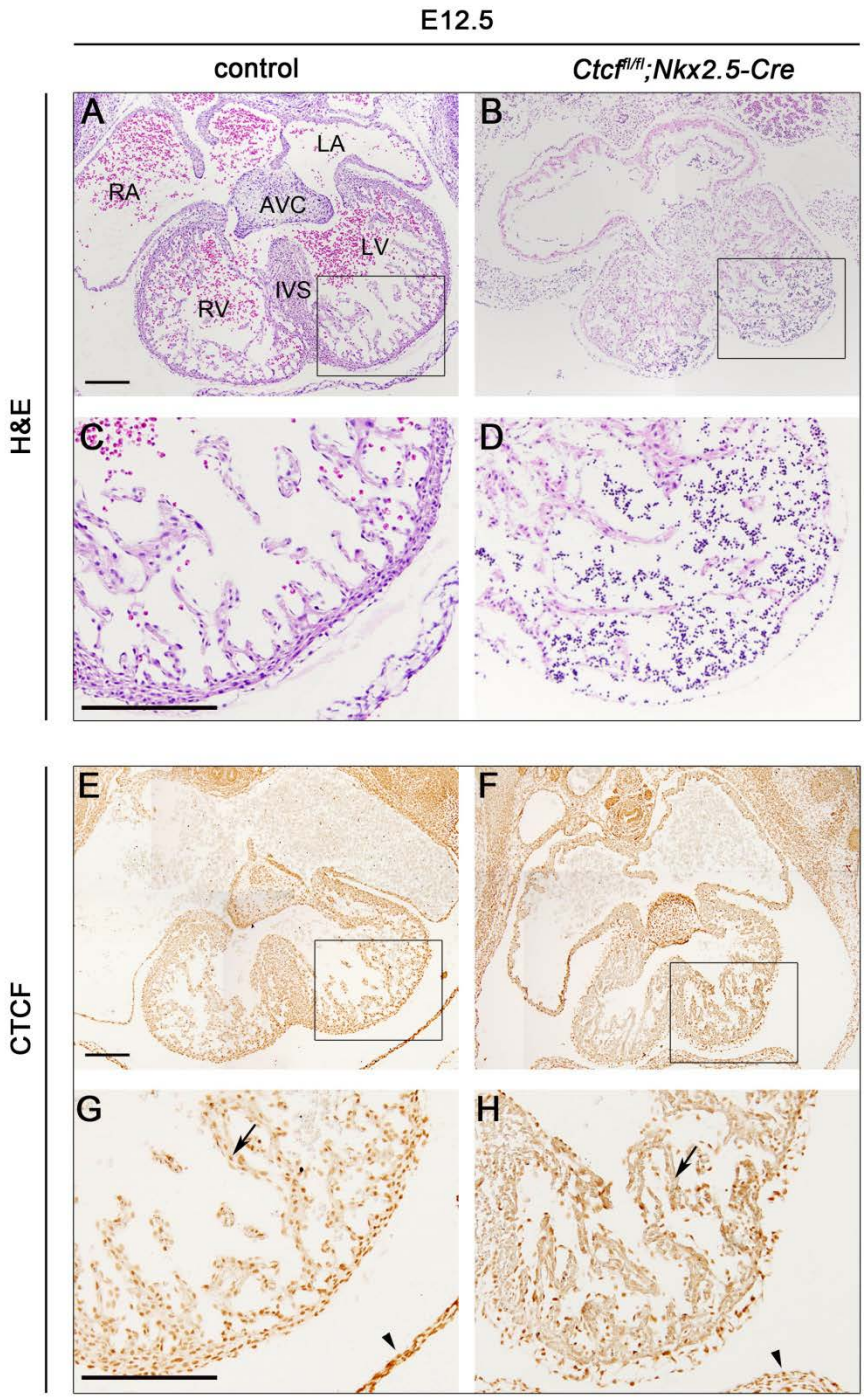

Fig. 14. *Ctcf* deletion in the heart causes severe cardiac malformations at E12.5.

A,B, H&E staining in transverse sections of control and *Ctcf* KO E12.5 hearts. *Ctcf* KO showed severe morphological defects. **C,D,** Closer view of the left ventricles of control and *Ctcf* KO hearts, respectively. **E,F,** CTCF immunostaining showed reduced CTCF expression in the *Ctcf* KO heart. This downregulation is better appreciated in **G** and **H**. Arrows show *Nkx2.5* territories and arrowheads show non-*Nkx2.5* territories, and point to the pericardium. Scale bars are 200 µm. RA: right atrium, RV: right ventricle, IVS: interventricular septum, AVC: atrioventricular canal, LA: left atrium, LV: left ventricle.

1.3. Cardiac *Ctcf* deletion does not alter apoptosis or cell proliferation

Previous studies in which *Ctcf* was deleted have shown that this deletion in some cases upregulates apoptosis (Hirayama et al. 2012; Soshnikova et al. 2010) and in others impedes cell proliferation (Heath et al. 2008; Ribeiro de Almeida et al. 2011). We therefore questioned whether these processes were altered in our model. To address this, we examined apoptosis by quantifying the number of TUNEL-positive nuclei, and we measured cell proliferation by quantifying PH3-positive nuclei in the whole heart. We chose to do this at E10.5 because this stage was the middle point between the onset of the conditional deletion (E7.5) and embryonic death (E13.5). Moreover, at this stage we could already detect CTCF downregulation (appendix E) and no overt morphological abnormalities. Immunostaining showed no apparent differences in the number of TUNEL-positive cells or PH3-positive nuclei between *Ctcf* KO and control hearts (Fig. 15 A,B, and D,E, respectively). These observations were corroborated by quantifying the number of cells, which were not different between the two samples. (Fig. 15 C,F). These results show that deletion of *Ctcf* in the heart does not alter apoptosis or cell proliferation at E10.5, although we cannot rule out that these processes are affected at later stages.

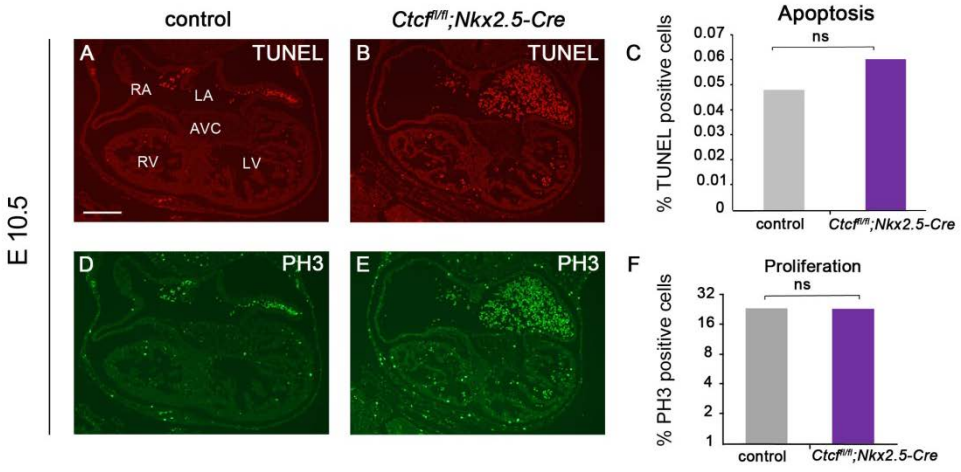

Fig. 15. *Ctcf* deletion in the heart does not alter apoptosis or proliferation.

A-C, TUNEL assays at E10.5 in control (**A**) and *Ctcf* KO hearts (**B**) revealed no differences between them **(C). D-F,** PH3 immunostaining in E10.5 control (**D**) and *Ctcf* KO hearts (**E**) also showed no difference between them **(F),** Scale bar is 200 µm. RA: right atrium, RV: right ventricle, LA: left atrium, LV: left ventricle, AVC: atrioventricular canal.

2. *Ctcf* is necessary for the correct expression of cardiac genes

To obtain a broader insight into the impact of deleting *Ctcf* in the developing heart, we decided to evaluate its effect on gene transcription. Thus, we performed RNA

sequencing (RNAseq) in embryonic hearts of the following genotypes: *i) Ctcf* KO (*Ctcf*$^{fl/fl}$*;Nkx2.5-Cre*), *ii)* controls (*Ctcf*$^{fl/+}$), and *iii)* heterozygous (*Ctcf*$^{fl/+}$*;Nkx2.5-Cre*) at E10.5, because of the reasons pointed above. After filtering the read counts of all samples and replicates, 14011 genes were considered as being expressed for further analysis. Principal component analysis (PCA) showed that samples for each condition grouped together and separately from the other conditions, indicating low variability within each condition and sufficient information to be differentiated between them (Fig. 16). We then made three comparisons: *i*) *Ctcf* KO versus control, *ii*) *Ctcf* KO versus heterozygotes, and *iii*) heterozygotes versus control. Genes were considered to be differentially expressed at an adjusted p-value <=0.05, and were sorted by the absolute log2 Fold Change (Appendix C).

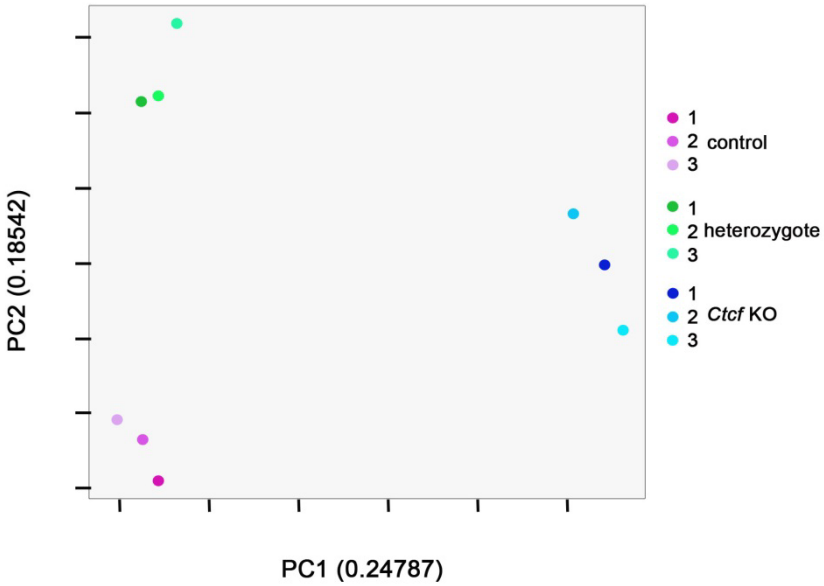

Fig. 16. Principal Component Analysis (PCA) of the RNAseq samples.

Graph shows how replicates of each sample group together, and how each sample group is separate from the other samples.

In the comparison *Ctcf* KO versus control, we obtained 2335 differentially expressed genes (DEG) from a total of 14011 genes considered in the analysis (Fig. 17A). Of these, 1126 were upregulated and1209 were downregulated. In the comparison *Ctcf* KO versus heterozygotes, we obtained 1989 DEG, from these 1022 were upregulated and 967 were downregulated (Fig. 17B). Finally, in the heterozygous versus control, we obtained only 24 DEG, 15 upregulated and 9 downregulated (Fig. 17C), indicating that both samples were very similar. After this, we compared the DEG from *Ctcf* versus control against the DEG from *Ctcf* KO versus heterozygous and found 1422 genes in common, representing 60 and 71% of each group of DEG respectively. These results

indicate that one copy of *Ctcf* is sufficient for proper transcription at E10.5, and at least in the heart, *Ctcf* does not act in a dosage dependent manner.

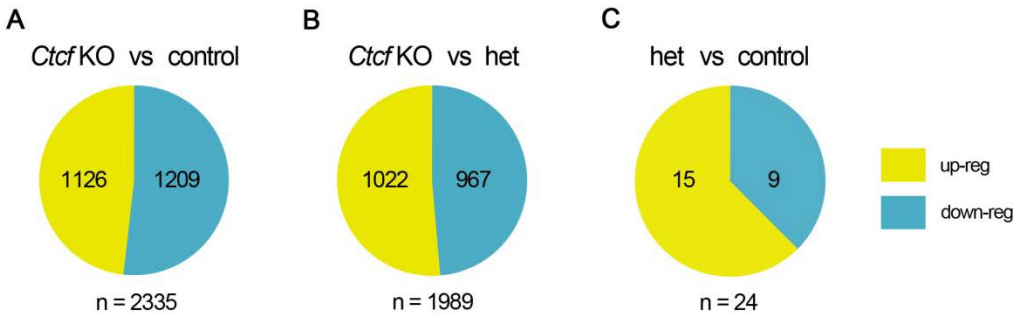

Fig. 17. RNAseq summary.

Pie charts showing the three comparisons of gene expression in embryonic hearts analyzed by RNAseq for *Ctcf* KO versus control (**A**), *Ctcf* KO versus heterozygous (**B**), and heterozygous versus control (**C**). The upregulated and the downregulated genes are depicted in yellow and blue, respectively. The numbers inside the circles denote the precise number of genes of each category in each comparison. n indicates the total number of genes differentially expressed in each comparison. het: heterozygous, up-reg: upregulated, down-reg: downregulated.

As a first approach to study the function of the DEG, we performed Gene Ontology (GO) analysis for Biological Processes terms using the DAVID functional annotation tool (*https://david.ncifcrf.gov*). In the comparison *Ctcf* KO versus control, we found enrichment of GO terms related to metabolism (such as tetrapyrrole biosynthetic process, cellular response to oxidative stress and cell cycle checkpoint, among others), and development (such as heart morphogenesis and heart development) (Fig. 18A and Table 16 appendix D). Subsequently, we divided DEG (of this comparison) into genes that were upregulated or downregulated in the mutant hearts compared to controls, and repeated the GO analysis with each group. Each data set was enriched for different GO terms. The GO terms related to heart development and function were present only in the downregulated dataset (Fig. 18B and Table 17 appendix D). Genes related to general developmental pathways such as NOTCH, BMPs, TGFbeta ERBb, and also several cardiac transcription factors were only present in the downregulated dataset (Fig. 19). Unlike the downregulated genes, the upregulated genes were enriched for GO terms related to general metabolism (Fig. 18C and Table 18 appendix D). GO analysis of the DEG from the *Ctcf* KO versus heterozygous showed a similar behavior as the *Ctcf* KO versus control. GO terms related to heart development and morphogenesis were present in all DEG and in the downregulated dataset (Tables 19-22 appendix D).

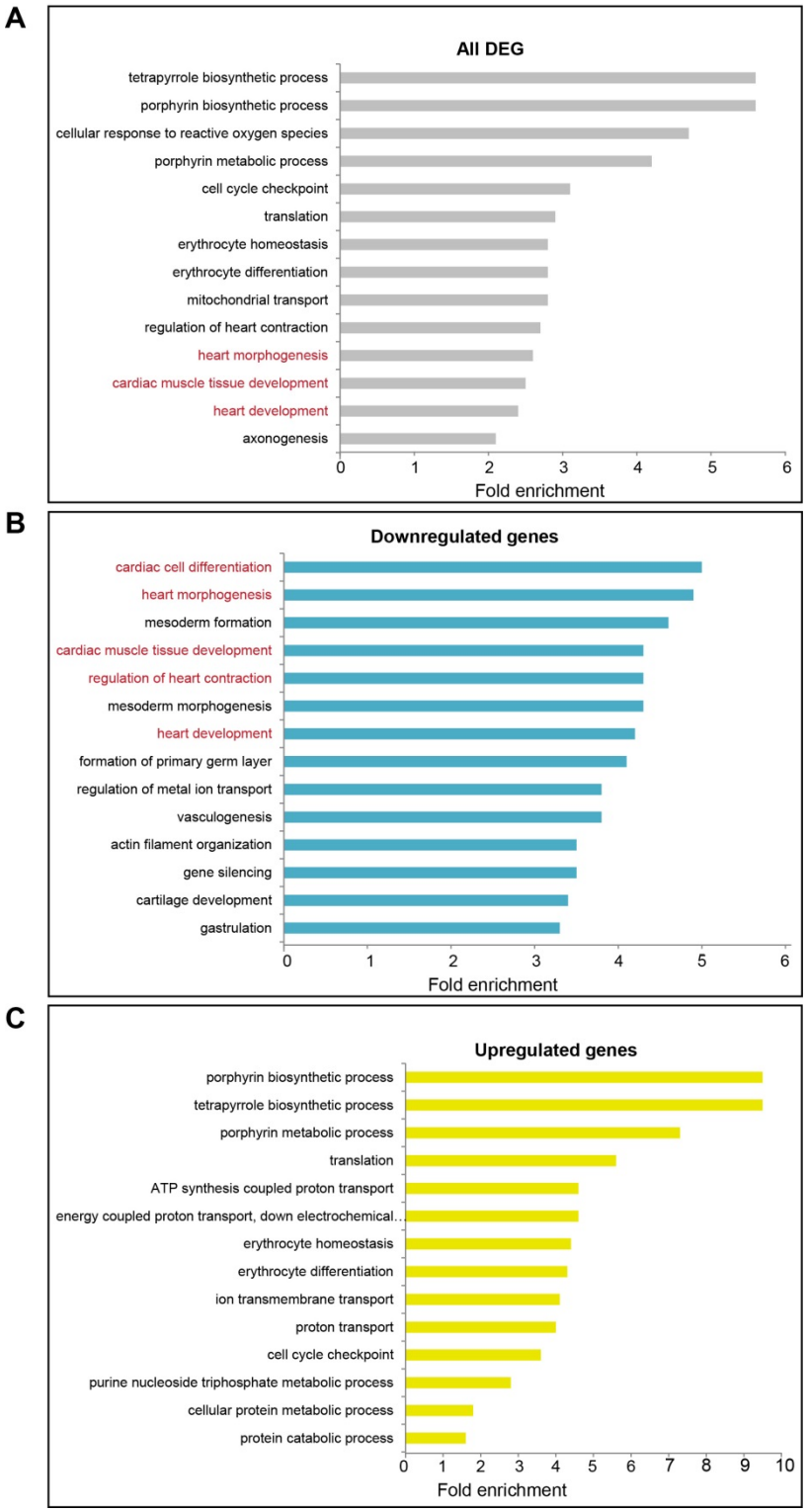

Fig. 18. Gene Ontology analysis for the *Ctcf* KO versus control comparison.

Each panel shows the top 15 GO terms in each data set, all DEG (**A**), downregulated genes (**B**), and upregulated genes (**C**). The GO terms are ordered from high- to low-fold change enrichment.

From all the DEG present in the *Ctcf* KO versus control comparison, we selected several cardiac transcription factors, such as *Nkx2*.5, *Hopx*, *Tbx20*, and *Pitx2*, also the cardiac chamber specific gene *Nppa* to evaluate their expression by *in situ* hybridization. Confirming the RNAseq data from the *Ctcf* KO versus control, we observed in some cases a general downregulation in gene expression, meaning that the downregulation was observed in the gene expression domain, in *Ctcf* KO hearts, for example *Nkx2*.5 (Fig. 10A), *Hopx* (Fig. 10B) and *Tbx20* (Fig.10C). Moreover, other genes showed reduced expression in specific territories, for example *Nppa*, which was absent in the anterior ventricular trabeculae (asterisks in Fig. 20D), and *Pitx2* in the right ventricle of KO hearts (arrows in Fig. 20E). *In situ* hybridization also confirmed strong upregulation of *Fgf13* upon *Ctcf* deletion (Fig. 20F), a fibroblast growth factor homolog that is suggested to contribute to heart disease (Wei et al. 2011), and this was among the most highly upregulated genes in the *Ctcf* KO as compared to the control.

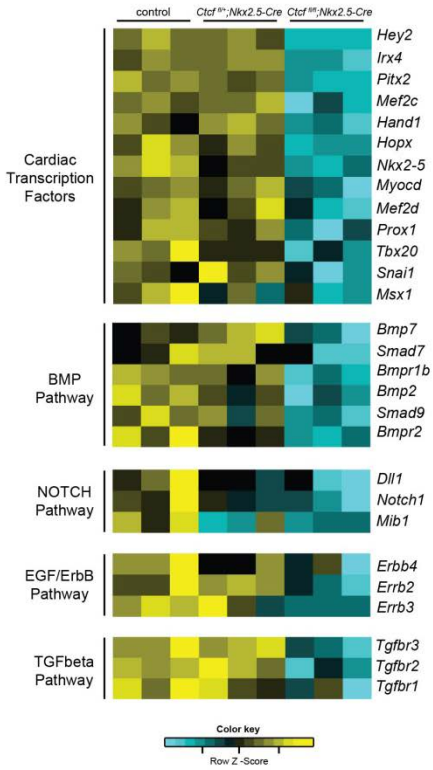

Fig. 19. *Ctcf* deletion alters cardiac developmental pathways.

Heat map showing the tendency of genes related to heart development (transcription factors and general developmental pathways such as BMP, NOTCH, EGF/ErbB and TGFbeta) to be downregulated in the in *Ctcf* KO hearts compared with controls.

3. *Ctcf* facilitates enhancer-promoter interactions in the embryonic heart

Previous studies have shown that there are several thousand *Ctcf* binding sites distributed genome-wide (H. Chen et al. 2012a; Shen et al. 2012). We examined the relationship between the DEG in the *Ctcf* KO versus control comparison and the presence of CTCF binding sites in their vicinity. To do this, we used the public available data (Shen et al. 2012) for CTCF Chip-seq in E14.5 and 8-week-old mouse hearts.

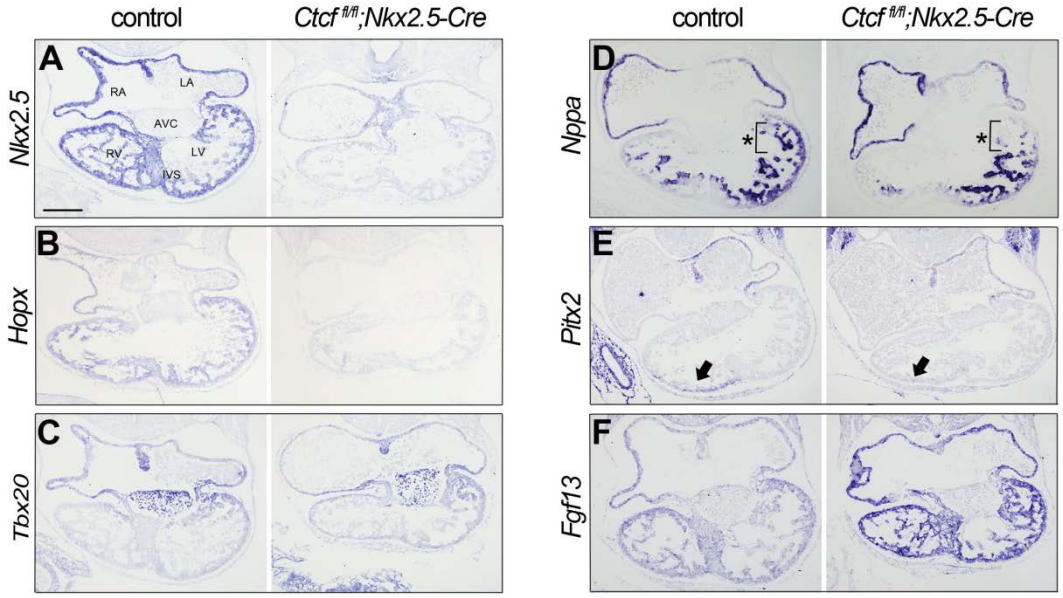

Fig. 20. *Ctcf* deletion alters heart developmental gene expression.

In situ hybridization shows that *Ctcf* deletion alters gene expression in the developing heart. It can cause general downregulation, as is the case for *Nkx2.5* (**A**), *Hopx* (**B**), or *Tbx20* (**C**); or downregulation in specific domains of expression, for example *Nppa* (**D**, asterisks show anterior ventricular trabeculae), *Pitx2* (**E**, arrows point to the right ventricle). *Fgf13* (**F**), is a fibroblast growth factor that is upregulated in *Ctcf* KO. RA: right atrium, RV: right ventricle, IVS: interventricular septum, AVC: atrioventricular canal, LA: left atrium, LV: left ventricle. Scale bar is 200 μm.

Furthermore, since we previously observed that upregulated and downregulated genes were enriched in different GO terms (Fig. 18), we decided to also evaluate them separately in this analysis. We used the genes present in the RNAseq that were not unchanged in the *Ctcf* KO versus control comparison (appendix C) as the background for the comparisons. Thus, in each dataset, we mapped the Transcription Start Site (TSS) of each gene, and used this position as a fixed reference point. We then utilized the available public data (Shen et al. 2012) to assess the distance from the TSS of each DEG to its closest CTCF binding site. We then compared these distances for the downregulated or upregulated genes versus those for the expressed but unchanged genes. We found that the upregulated and downregulated genes were closer to a

CTCF binding site than genes whose expression did not change upon *Ctcf* deletion (Fig. 21A). Next, we questioned the relationship between the same elements (TSS of the DEG and the CTCF binding sites) when examined in a delimited genomic window of 10 or 20 kb surrounding the TSS. We assessed the DEG separately (upregulated and downregulated versus expressed but unchanged in the *Ctcf* KO versus control comparison) and queried the number of DEG that had a CTCF binding site inside the delimited genomic window. We found that both, downregulated and upregulated genes presented a significantly higher proportion of genes that had a CTCF binding site within 10 or 20 kb genomic windows when compared with the expressed but unchanged genes in the *Ctcf* KO versus control comparison (Fig 21B).

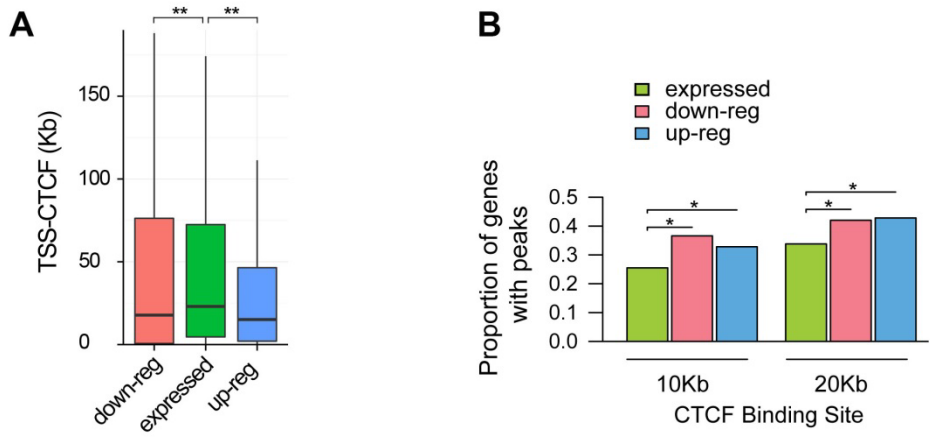

Fig. 21. Distance from the DEG from *Ctcf* KO versus control comparison to the nearest CTCF binding site.

A, Differentially expressed genes (DEG) in the *Ctcf* versus control comparison, either upregulated (blue) or downregulated (pink), lie closer to a CTCF binding site than the expressed but unchanged genes (green). **, p<0.0005; Mann-Whitney test for DEGs vs expressed *P* = 2.475797e-13. **B,** This tendency is also appreciated in genomic windows of 10 and 20 kb. *, p<0.01; Proportions test. *P* values are listed as follows: in the 10 Kb window, down-reg versus expressed *P* = 1.54E-16, and up-reg versus expressed *P* = 0.000000121; in the 20 Kb window, down-reg versus expressed *P* = 1.42E-08, and up-reg versus expressed *P* = 1.79E-09.

Shen and colleagues (2012) used the chromatin histone modification H3K4me1 together with p300 promoter-distal binding sites to train an enhancer prediction tool, and predict heart enhancers for E14.5 and 8-week adult stages (Shen et al. 2012). We used these predicted heart enhancers to examine the relationship between them and the DEG in the *Ctcf* KO heart versus the control. Again, we evaluated this separately in the downregulated and upregulated genes. As in the analysis of the relationship between CTCF binding sites and DEG (Fig. 21A), we mapped the TSS of each gene and used this position as a fixed reference point. We assessed the distance from each

DEG to its closest predicted heart enhancer, and then compared these distances from the downregulated or upregulated genes versus those for the expressed but unchanged genes in the *Ctcf* KO versus control comparison. We found that both datasets were significantly closer to a predicted heart enhancer than genes whose expression did not change upon *Ctcf* deletion (Fig. 22A). After this, we questioned the relationship between the same elements (TSS of the DEG and the predicted enhancers) when examined in a delimited genomic window of 10 and 20 kb. We assessed the DEG separately (upregulated and downregulated versus expressed but unchanged in the *Ctcf* KO versus control comparison) and queried the number of DEG that had a predicted heart enhancer inside the delimited genomic windows. We found that only the downregulated genes presented a significantly higher proportion of genes that had a predicted heart enhancer within 10 or 20 kb genomic windows, when compared with the control (Fig 22B). Although DEG genes are closer to a heart enhancer, the distance between the upregulated genes to a heart enhancer is similar to the expressed but unchanged genes. Thus, when we restricted the genomic windows to 10 and 20 kb, only downregulated genes are significantly closer to a heart enhancer.

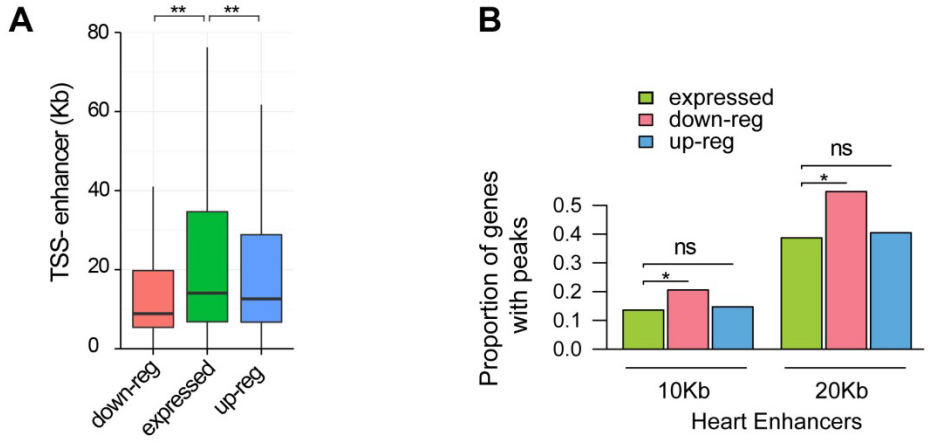

Fig. 22. Distance from the DEG from *Ctcf* KO versus control comparison to the nearest heart enhancer.

A, Genes downregulated (pink) and upregulated (blue) in the *Ctcf* KO versus control comparison lie closer to a previously described heart enhancer than unchanged (green) genes. **, p<0.0005; Mann-Whitney test. Exact *P* values are listed as follows: down-reg versus expressed *P* = 9.209248e-31, up-reg versus expressed *P* = 0.002601427. **B**, In short genomic windows of 10 and 20 kb only downregulated genes lie closer to a heart enhancer. *, p<0.01; Proportions test. Exact *P* values are listed as follows: in the 10 Kb window, down-reg versus expressed *P* = 3.83E-11, and up-reg versus expressed *P* = 0.347; in the 20 Kb window, down-reg versus expressed *P* = 1.8E-27, and up-reg versus expressed *P* = 0.235. n.s.: non-significant.

We next questioned whether there were other differences between the upregulated and downregulated genes from the *Ctcf* KO versus control comparison. We therefore regarded the gene distribution, i.e., how far away or close were the DEG relative to other genes in the genome. To do this, we again used the TSS of each gene as a fixed position, only this time we measured the distance (in kb) between the TSS of the DEG and the TSS of the immediately neighboring upstream and downstream genes. We summed both distances for each gene and obtained a single value, the total distance (expressed in kb) for each gene. We then compared this value (the total distance) between upregulated and downregulated genes. We again observed a difference between these two datasets. The downregulated genes presented a larger distance to neighboring genes than the upregulated genes (Fig. 23).

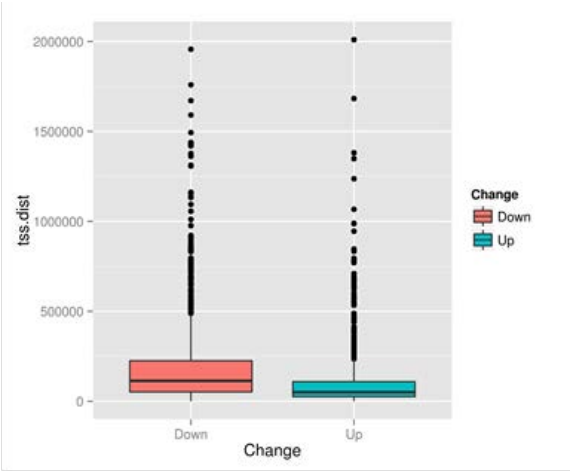

Fig. 23. Distance between the DEG from the *Ctcf* KO versus control comparison with their neighboring genes.

4. *Ctcf* deletion affects the expression of tandemly arranged genes

Once we had established a global view of the consequences for gene transcription of *Ctcf* deletion in the embryonic heart, we sought to further explore the role of *Ctcf* in gene regulation in heart development by focusing on specific examples. *Ctcf* can act as an insulator, and the CTCF binding sites can act by delimiting regulatory landscapes (Narendra et al. 2015; Yang and Corces 2012). One way to test the role of *Ctcf* as an insulator is to examine the expression pattern (and levels) of gene pairs that are in tandem (next to each other), differentially expressed in control conditions, and that have at least one CTCF binding site between them that could explain the differential expression. Several gene pairs met these criteria. We selected the following four gene pairs to examine the role of *Ctcf* as an insulator, and analyzed their expression by RNA

in situ hybridization: *Tnnt1-Tnni3* (Fig. 24A), *Tnni2-Tnnt3* (Fig. 15A), *Tbx2-Tbx4* (Fig. 16A) and *Tbx3-Tbx5* (Fig. 27A).

The troponin complex plays a role in the regulation of striated muscle contraction. It has three subunits: troponin I (*Tnni*), Troponin T (*Tnnt*) and Troponin C (*Tnnc*), they are the inhibitory, tropomyosin-binding and calcium-binding subunit respectively (Katrukha 2013). Each subunit has homologous that are tissue specific. We focus here in four troponin genes. *Tnni3* is cardiac specific and is expressed in all myocardium (Christoffels et al. 2006; Christoffels et al. 2009; Zhu et al. 1995). *Tnnt1* is transiently expressed in the developing heart, but not exclusively. In the heart it is expressed at lower levels in comparison to *Tnni3*, and *Tnnt1* expression pattern is not uniform within the myocardium, it is not expressed at the same levels in all myocardium (Barton et al. 2004; Q. Wang et al. 2001). There is one study that detects *Tnni2* in the developing heart (Zhu et al. 1995). It is expressed at lower levels in comparison to *Tnni3*. *Tnni2* expression pattern is not uniform and is restricted to the compact myocardium (Zhu et al. 1995). *Tnnt3* is expressed in non-cardiac striated muscle (Q. Wang et al. 2001). *Tnnt1*, *Tnni2* and *Tnnt3* are upregulated in the heart in cardiomyopathies (Huang et al. 2008; Katrukha 2013; Petchey et al. 2014).

The troponin gene pairs analyzed here presented significantly changed expression levels in the *Ctcf* KO versus control comparison. Interestingly, in both gene pairs, one was downregulated and the other one was upregulated (*Tnnt1-Tnni3* and *Tnni2-Tnnt3*, respectively).

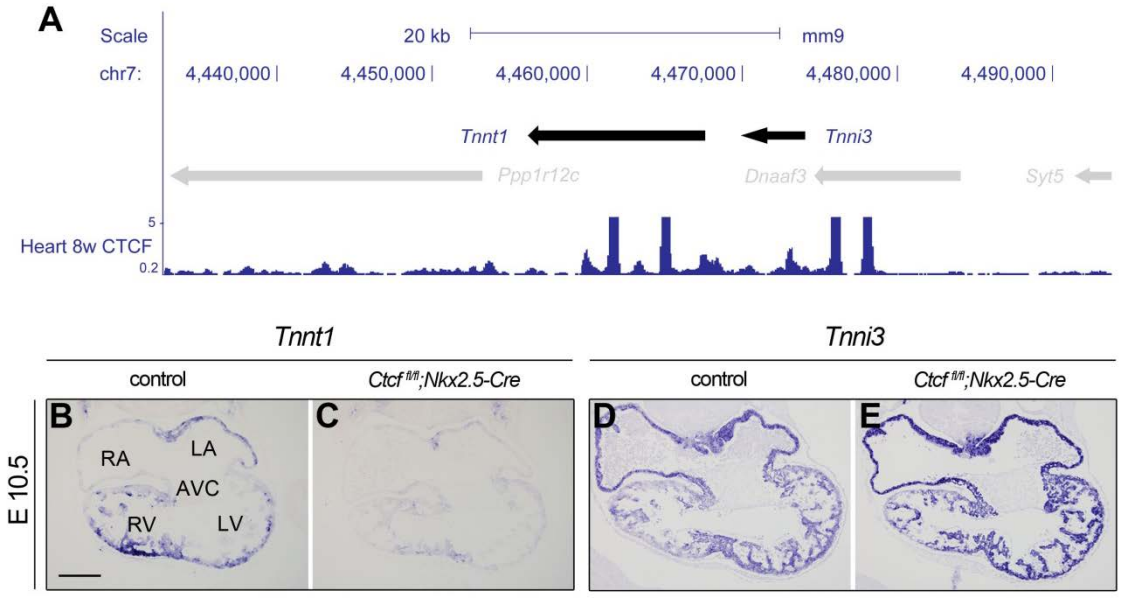

Fig. 24. *Tnnt1* and *Tnni3* expression upon *Ctcf* deletion.

A, Genomic region containing *Tnnt1* and *Tnni3* (mm9, chr7:4,433,080- 4,493,651) together with the distribution of CTCF binding sites in the heart. Troponin genes are highlighted in black. **B-E**, *In situ* hybridization in sections of E10.5 control and KO hearts for *Tnnt1* (**B,C**) and *Tnni3* (**D,E**). RA: right atrium, RV: right ventricle, AVC: atrioventricular canal, LA: left atrium, LV: left ventricle. Scale bar is 200 µm.

RNA *in situ* hybridization at E10.5 showed that *Tnnt1* was expressed mainly in the compact myocardium of the ventricle and in the left atrium (Fig. 24B). We found that indeed *Tnni3* was expressed in all the compact myocardium, the trabeculae and in both atria (Fig. 24D). *Tnnt1* expression was drastically decreased in the *Ctcf* KO (Fig. 24C), and *Tnni3* expression was modestly increased (Fig. 24E).

We found that in the control, *Tnni2* was expressed in the ventricles and atria (Fig. 25B) and *Tnnt3* was expressed as a few scattered dots only in the atria (Fig. 25D). In the *Ctcf* KO, *Tnni2* expression was dramatically decreased (Fig. 25C), whereas *Tnnt3* expression was increased only in the atria (Fig. 25E).

Results from RNA *in situ* hybridization of the troponin genes here analyzed fitted well with those from the RNAseq analysis for the *Ctcf* KO versus control comparison. Given that these genes presented divergent expression patterns, we hypothesized that they belong to different regulatory domains. Moreover, CTCF can help to delimitate regulatory domains (Narendra et al. 2015; Yang and Corces 2012), in other words to form and/or maintain boundaries. If the regulatory domains were separated by CTCF, we would expect ectopic expression patterns of at least one member of the gene pairs. This could be explained because without the insulator, inadequate regulatory elements could act upon the inappropriate gene. We only found up or downregulation of genes in their endogenous domains of expression. *Ctcf* loss does influences expression of the four troponin genes analyzed here, but it does not act by insulating the regulatory elements in the troponins gene pairs.

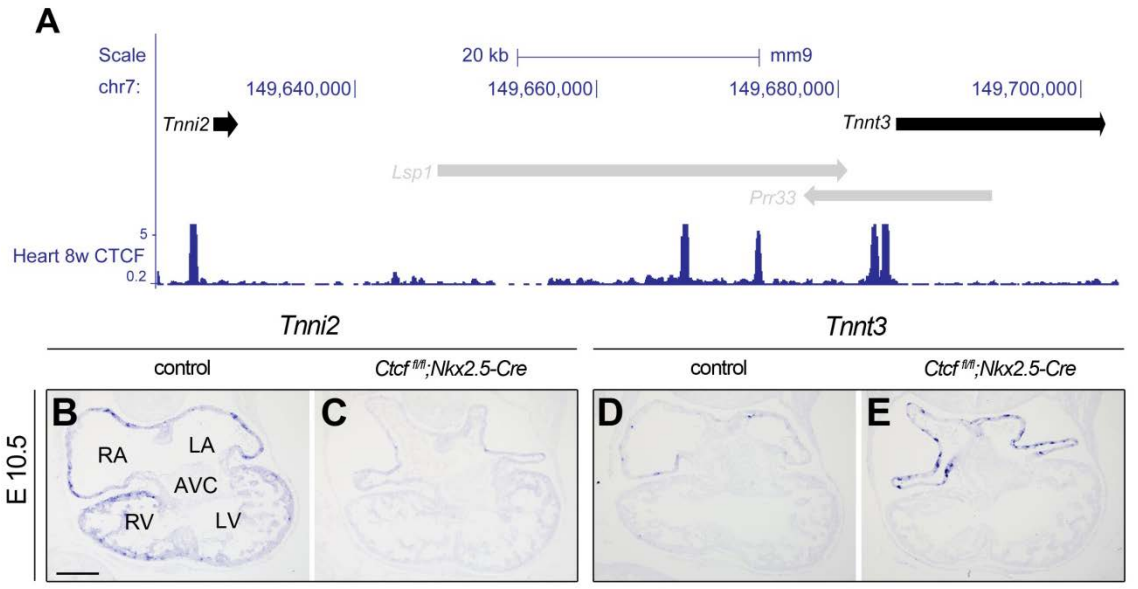

Fig. 25. *Tnni2* and *Tnnt3* expression upon *Ctcf* deletion.

A, Genomic region containing *Tnni2* and *Tnnt3* (mm9, chr7:149,623,608-149,703,181) together with the distribution of CTCF binding sites in the heart. Troponin genes are highlighted in black. **B-E,** *In situ* hybridization in sections of E10.5 control and KO hearts for *Tnni2* (**B,C**) and *Tnnt3* (**D,E**). RA: right atrium, RV: right ventricle, AVC: atrioventricular canal, LA: left atrium, LV: left ventricle. Scale bar is 200 µm.

The T-box transcription factors are a group of genes, at least 18, that play a role in heart development (Clowes et al. 2014). We focus here in four genes that are arranged in tandem by pairs (*Tbx2-Tbx4*) and (*Tbx3-Tbx5*). This genomic arrangement is conserved in evolution (Horton et al. 2008). *Tbx5* is expressed in both atria and left ventricle, and it activates chamber myocardial gene program (Greulich et al. 2011). Haploinsufficiency of this gene causes Holt-Oram syndrome (Bruneau et al. 1999). *Tbx2* and *Tbx3* represses chamber specific genes and block chamber formation. They are expressed in non-chamber territories in the heart, like atrioventricular canal (*Tbx2* and *Tbx3*) and out flow tract (*Tbx2*) (Clowes et al. 2014; Hoogaars et al. 2004). *Tbx4* is expressed in the developing heart, but no role in heart morphogenesis has been so far described (Chapman et al. 1996; Krause et al. 2004).

These four genes were not differentially expressed in the *Ctcf* KO versus control comparison. Still, we wished to analyze if subtle changes in expression could be detected by RNA *in situ* hybridization. To analyze *Tbx* gene expression, we performed RNA *in situ* hybridization at two developmental stages, E9.5 and E11.5, to evaluate whether changes in expression could be detected at either or both stages.

In the control, *Tbx2* was expressed in the atrioventricular canal (Fig. 26B-D and (Stennard and Harvey 2005)). No expression changes were detected in KO hearts at

E9.5 (Fig. 26B-D), and only subtle ectopic expression in the right ventricle at E11.5 was detected (arrowheads, Fig. 26D-E). *Tbx4* is associated with limb development, however, it has also been detected in the heart by *in situ* hybridization (Chapman et al. 1996) and it was also detected in our RNAseq (appendix C). We were only able to detect extremely subtle *Tbx4* expression in the heart in controls, with no obvious changes in KOs (Fig. 26F-I).

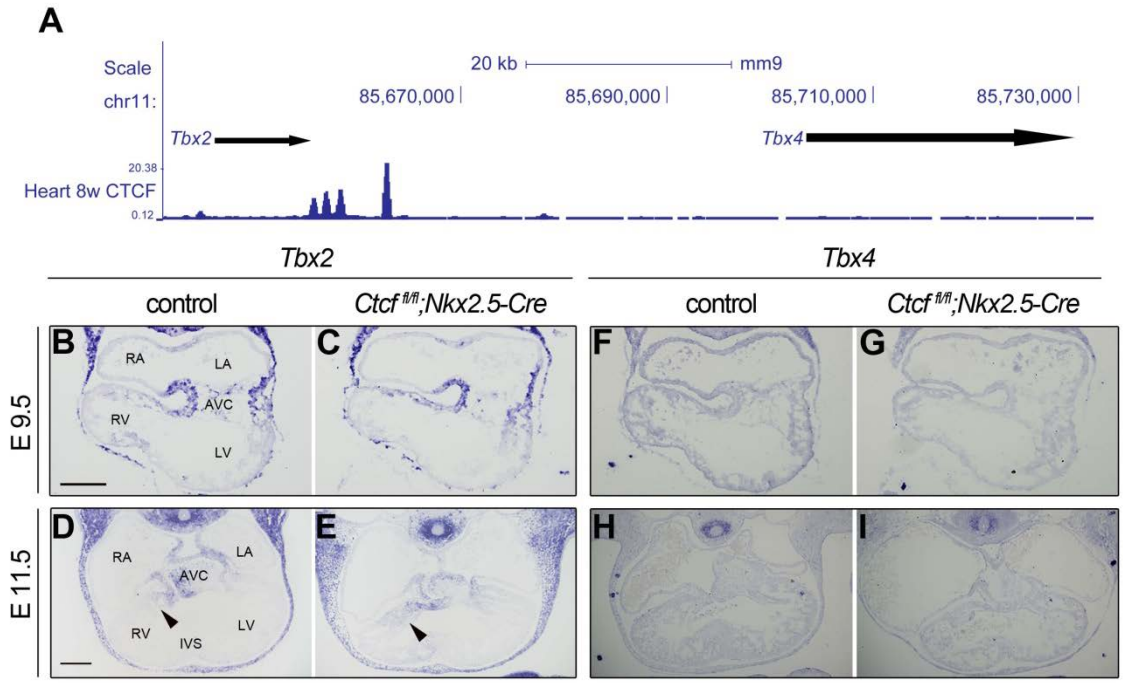

Fig. 26. *Tbx2* and *Tbx4* expression upon *Ctcf* deletion.

A, Genomic region containing *Tbx2* and *Tbx4* (mm9, chr11:85,641,154-85,731,359) together with the distribution of CTCF binding sites in the heart. *Tbx* genes are highlighted in black. **B-I**, *In situ* hybridization in sections of control and KO hearts for *Tbx2* at E9.5 (**B,C**) and E11.5 (**D,E**). Arowheads at **D** and **E** show subtle *Tbx2* upregulation; and *Tbx4* at E9.5 (**F,G**) and E11.5 (**H,I**). RA: right atrium, RV: right ventricle, IVS: interventricular septum, AVC: atrioventricular canal, LA: left atrium, LV: left ventricle. Scale bar is 200 µm.

Finally, we examined the expression of the *Tbx3-Tbx5* gene pair. *Tbx3* was expressed in the atrioventricular canal in control mice, and we found that its expression was unaltered upon *Ctcf* deletion at any stage (Fig. 27B-E). In control hearts, *Tbx5* was expressed in both atria and also in the left ventricle. In *Ctcf* KO hearts, its expression was decreased only at E11.5, with no obvious changes in levels or pattern of expression at E9.5 (Fig. 27F-I). Similar to our results with the troponin genes, we failed to detect ectopic expression in the *Ctcf* KO hearts of any of the *Tbx* genes analyzed. While *Tbxs* genes also presented characteristics that make them appealing candidates to test the role of *Ctcf* as an insulator, again we found that *Ctcf* seems not to act as an

insulator to separate the adjacent regulatory domains of the Tbx gene pairs here analyzed.

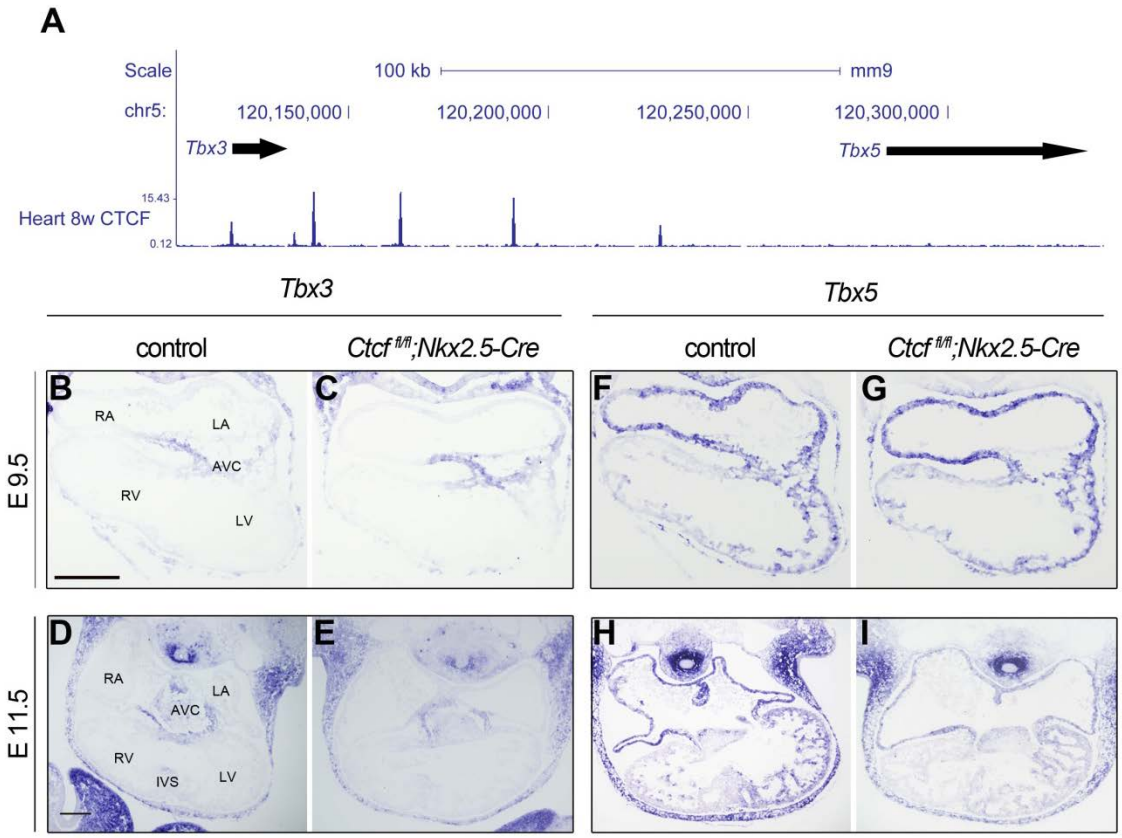

Fig. 27. *Tbx3* and *Tbx5* expression upon *Ctcf* deletion.

A, Genomic region containing *Tbx3* and *Tbx5* (mm9, chr5:120,107,016-120,338,997) together with the distribution of CTCF binding sites in the heart. *Tbx* genes are highlighted in black. **B-I,** *In situ* hybridization in sections of control and KO hearts for *Tbx3* at E9.5 (**B,C**) and E11.5 (**D,E**); and *Tbx5* at E9.5 (**F,G**) and E11.5 (**H,I**). RA: right atrium, RV: right ventricle, AVC: atrioventricular canal, LA: left atrium, LV: left ventricle. Scale bar is 200 µm.

5. *Ctcf* alters the expression of genes in the IrxA cluster

To further characterize the role of *Ctcf* in gene regulation, we focused our attention on *Irx4*, a chamber identity cardiac transcription factor that belongs to the *IrxA* cluster (Bao et al. 1999; Gomez-Skarmeta and Modolell 2002). This cluster includes three genes: *Irx1*, *Irx2* and *Irx4*. All genes from this cluster are expressed in the heart (Kim et al. 2012). Only *Irx4* was differentially expressed in *Ctcf* KO versus control heart comparisons, and it was downregulated (Fig. 28). *Irx1* and *Irx2* showed no significant changes, however they presented the tendency to be upregulated (Fig. 28). We thus performed RNA *in situ* hybridization of *Irx1*, *Irx2* and *Irx4* genes at E9.5 and E11.5. *Irx4* is strongly expressed in the ventricles, whereas *Irx1*/*Irx2* are expressed in the interventricular septum of the developing heart (Kim et al. 2012). At E9.5, *Irx1* and *Irx2*

showed a subtle expansion of their expression pattern in the ventricles (arrows in Fig. 29B,C), whereas *Irx4* showed no changes in expression levels at this stage (Fig. 29D). At E11.5, *Irx1* RNA *in situ* hybridization also showed a subtle expansion in its expression pattern (Fig. 29E,H); *Irx2* did not present any changes in its expression pattern (Fig. 29F), and *Irx4* showed a dramatic decrease in expression levels (Fig. 29G,I).

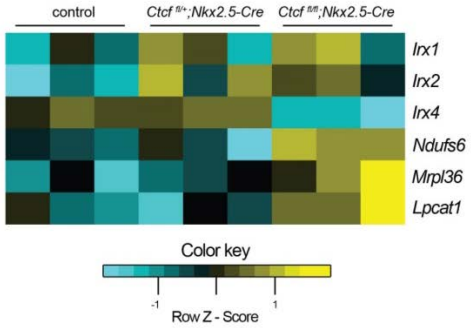

Fig. 28. IrxA cluster dysregulation upon *Ctcf* deletion.

Irx4 is the only member of the IrxA cluster found to be downregulated. The other members (*Irx1* and *Irx2*) do not change, and their three immediate neighbors (*Ndufs6*, *Mrpl36* and *Lpcat1*) are upregulated.

Further inspection of the RNAseq data showed that the three genes telomeric to the IrxA cluster (Fig. 30A), *Ndufs6*, *Mrlp36* and *Lpcat1*, were upregulated (Fig. 28). We performed RNA *in situ* hybridization of the first two genes, *Ndufs6* and *Mrlp36*. Only *Ndufs6* has been detected in the heart by RNA *in situ* hybridization (Smith et al. 2014). Our results showed that both genes were highly expressed in the whole heart and in the embryo (Fig. 31B,D), and only *Ndufs6* (Fig. 30C,E) presented a subtle upregulation in KO embryos by *in situ*.

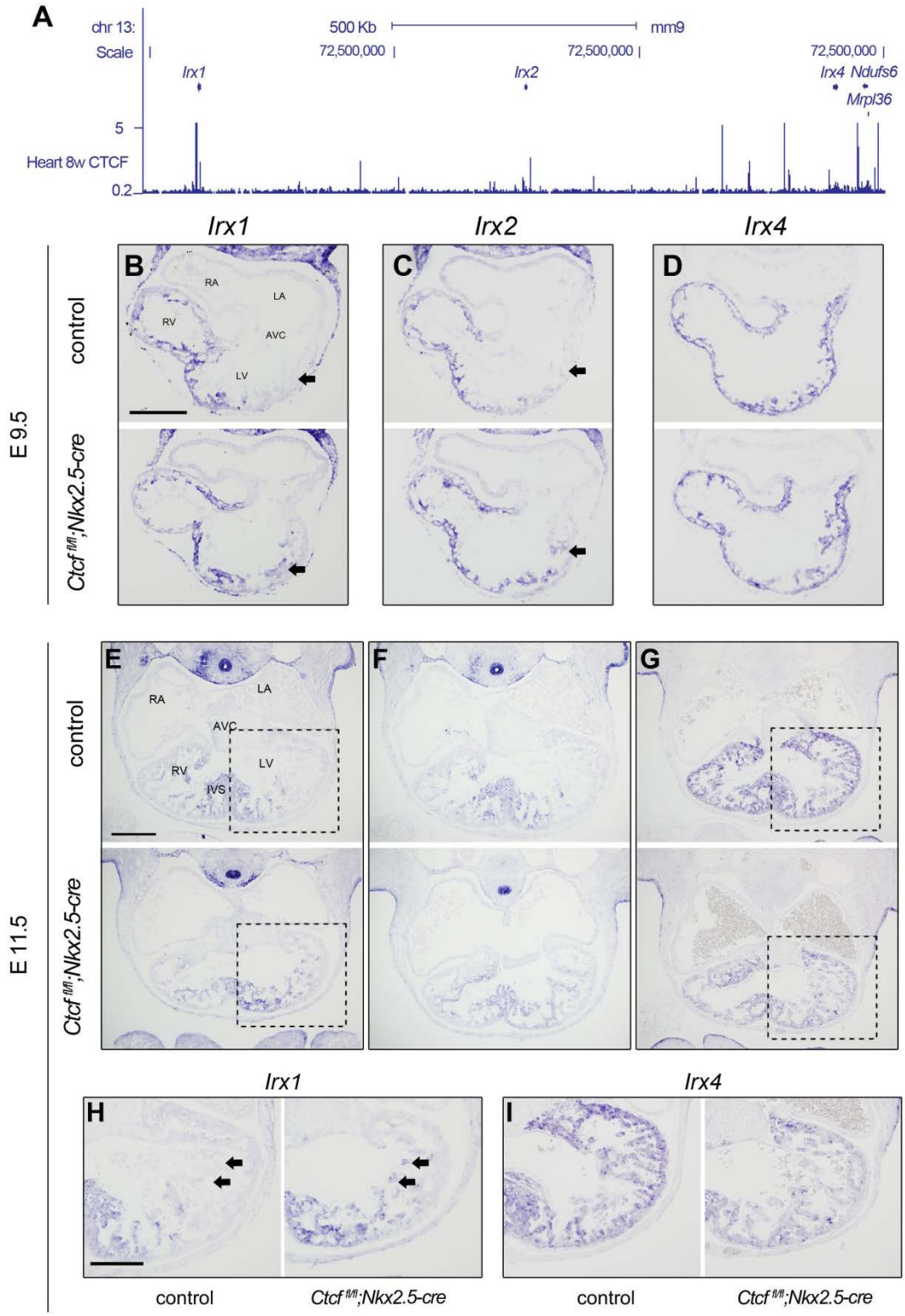

Fig. 29. IrxA cluster altered gene expression at E9.5 upon *Ctcf* deletion

A, Genomic region containing the IrxA cluster (mm9, chr13:71,984,688-73,504,230) together with the distribution of CTCF binding sites in the heart. **B-D**. *In situ* hybridization on sections of E9.5 control and KO hearts for *Irx1* (**B**), *Irx2* (**C**); and *Irx4* (**D**) Arrows in **B** and **C** show ectopic expression of *Irx1* and *Irx2* respectively. **E-I,** *In situ* hybridization in sections of E11.5 control and KO hearts for *Irx1* (**E**), *Irx2* (**F**); and *Irx4* (**G**). Close up views show the ectopic expression of

Irx1 (**H**, arrows) and downregulation of *Irx4* (**I**) RA: right atrium, RV: right ventricle, IVS: interventricular septum, AVC: atrioventricular canal, LA: left atrium, LV: left ventricle. Scale bar is 200 µm.

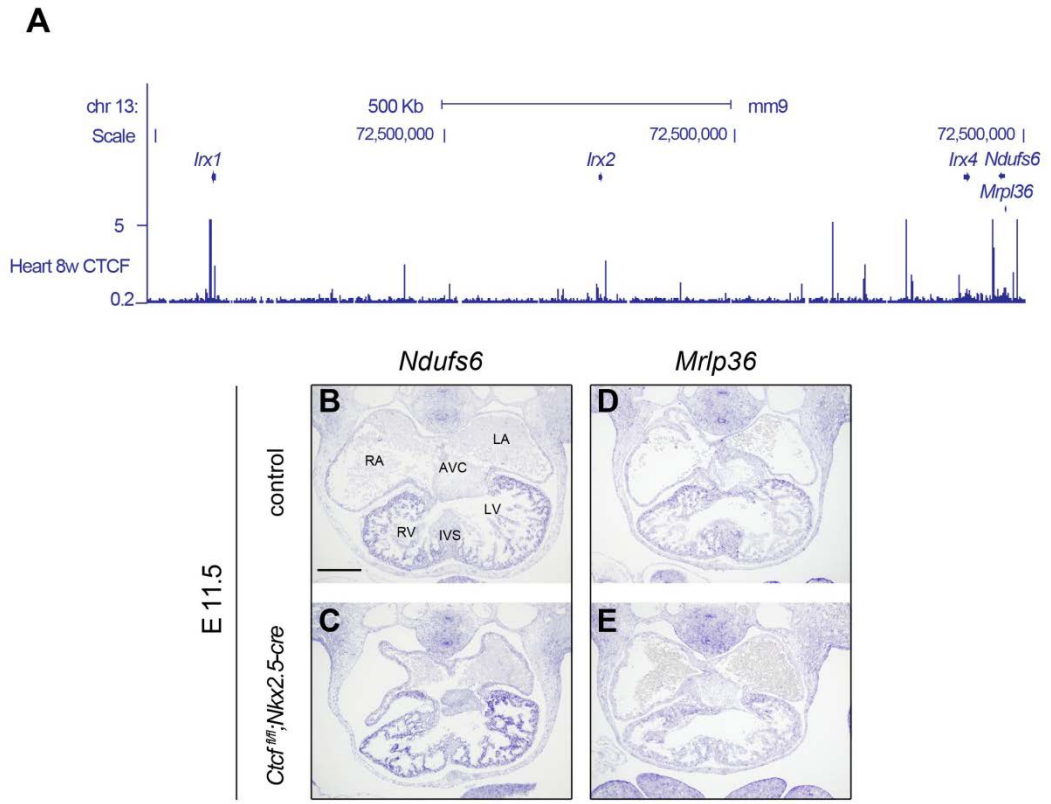

Fig. 30. Neighbors of the IrxA cluster with altered gene expression at E11.5 upon *Ctcf* deletion.

A. Genomic region containing the IrxA cluster and neighboring genes *Ndufs6 and Mrpl36* (mm9, chr13:71,984,688-73,504,230) together with the distribution of CTCF binding sites in the heart. **B-E**. *In situ* hybridization in sections of E11.5 control (**B**,**D**) and KO hearts (**C**,**E**) for *Ndufs6* and *Mrlp36* respectively. RA: right atrium, RV: right ventricle, IVS: interventricular septum, AVC: atrioventricular canal, LA: left atrium, LV: left ventricle. Scale bar is 200 µm.

The IrxA cluster has a homologous cluster, IrxB, and previous evidence from zebrafish has shown that CTCF might have a role in separating the first two genes, *Irx3* and *Irx5,* from the most telomeric gene of the cluster, *Irx6* (Tena et al. 2011). *Irx3* is expressed in the trabecular region of the ventricles, whereas *Irx5* and *Irx6* are expressed in the endocardium of the ventricles and atria (Kim et al. 2012). Although none of these genes were differentially expressed in the RNAseq analysis, we performed RNA *in situ* hybridization of the three genes at E11.5, which was the stage at which more dramatic changes were found in the expression of IrxA cluster genes in the *Ctcf* KO hearts. However, no member of the cluster presented a change in expression upon *Ctcf* loss at E11.5 (Fig. 31).

6. The three dimensional structure of the IrxA cluster

The Irx genes have a conserved genomic organization that occupies around 1 Mb (Gomez-Skarmeta and Modolell 2002). Previous work has shown that this conserved genomic organization is related to Irx gene regulation since the large genomic region in each cluster, with only three genes, harbors enhancers that are possibly shared among genes in the clusters (de la Calle-Mustienes et al. 2005; Tena et al. 2011). Moreover, 3C experiments in the IrxA cluster showed that physical interactions occur between several enhancers and the *Irx1* and *Irx2* promoters, and that these interactions are tissue specific (Tena et al. 2011). This raised the possibility that not all the members of the cluster share the same regulatory elements. Besides enhancer-promoter interactions, promoter-promoter interactions between *Irx1* and *Irx2* were also found, and in a tissue independent manner. Moreover, the *Irx1/Irx2* promoter interaction was conserved among vertebrates and it was suggested to be CTCF-dependent (Tena et al. 2011).

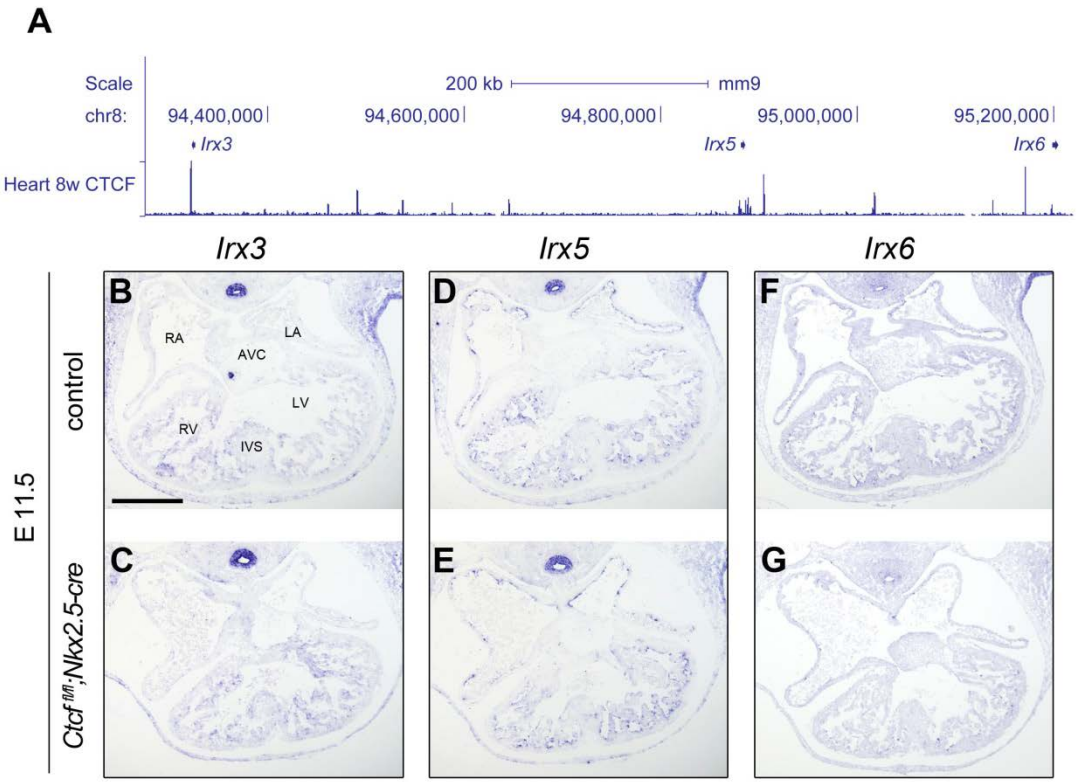

Fig. 31. IrxB cluster gene expression at E11.5 upon *Ctcf* deletion.

A, Genomic region containing the IrxB cluster (mm9, chr8:94,275,320-95,219,217) together with the distribution of CTCF binding sites in the heart. **B-G,** *In situ* hybridization in sections of E11.5 control and KO hearts for *Irx3* (**B,C**) *Irx5* (**D,E**), and *Irx6* (**F,G**). RA: right atrium, RV: right ventricle, IVS: interventricular septum, AVC: atrioventricular canal, LA: left atrium, LV: left ventricle. Scale bar is 200 μm

However, the lack of interaction between evolutionarily conserved enhancers from the IrxA cluster and *Irx4* (Tena et al. 2011) remains an unresolved issue. To get further insight into the 3D structure of the IrxA cluster, we carried out 4C-seq (Splinter et al. 2012; van de Werken et al. 2012a) using the promoters of *Irx1*, *Irx2, Irx4* and the first downstream gene of the cluster, *Ndufs6*, as viewpoints.

We observed that the IrxA cluster was divided into two interacting regions (Fig. 32). The centromeric part of the cluster contains *Irx1* and *Irx2* genes, and they presented a very similar interaction pattern that diminished at a similar point. *Irx4* is found in the telomeric part of the cluster. *Irx4* and *Ndufs6* had similar interaction patterns. Also, the *Irx1* and *Irx2* interacting region is larger than the region containing *Irx4* and *Ndufs6*. The division of the cluster fits with the previous suggestion that the members of the cluster could share some enhancers because they have few partially overlapping expression territories (Cavodeassi et al. 2001; Gomez-Skarmeta and Modolell 2002). However, *Irx1* and *Irx2* share more expression territories. Thus it is highly likely that these two genes share more regulatory elements between themselves that either one with *Irx4*. Our results support this possibility as we can see in the 4C-seq profile that these two genes present similar interacting profile.

7. Loss of CTCF disrupts the chromatin structure of the IrxA cluster

Given the observation that CTCF loss downregulated *Irx4* and altered the expression of other genes in the cluster (Fig. 29), we next questioned whether *Ctcf* deletion also altered the 3D structure of the IrxA cluster. To address this, we carried out 4C-seq in *Ctcf* KO hearts using the same four viewpoints used in the control hearts: *Irx1*, *Irx2, Irx4* and *Ndufs6*. We observed that the 3D structure of *Irx1*, *Irx2* and *Ndufs6* (Fig. 33) was basically preserved upon *Ctcf* loss, and that only *Irx4* presented significant changes in interactions between KO and control hearts (Fig. 34).

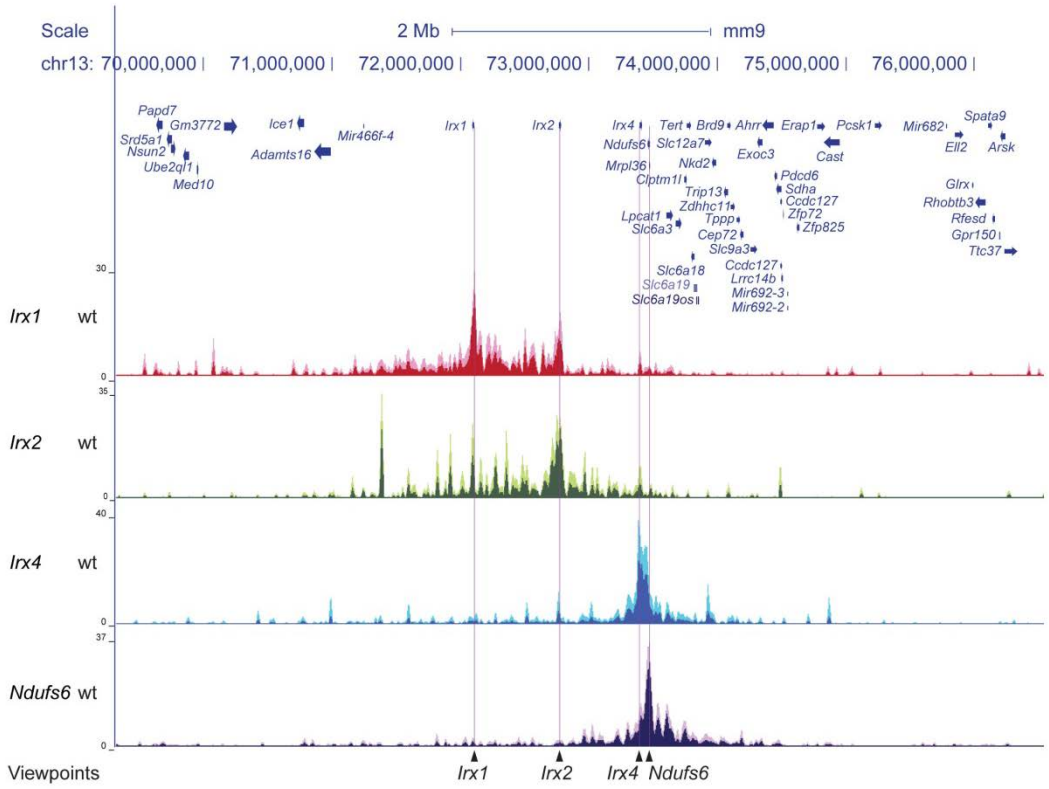

Fig. 32. The IrxA cluster is divided into two distinct interaction domains.

Chromatin interactions in the IrxA cluster in E11.5 control (wt) hearts revealed by 4C-seq using viewpoints in the promoters of *Irx1*, *Irx2*, *Irx4* and *Ndufs6* (vertical purple lines) (mm9; chr13:69,320,289-76,541,363). n≥3 for each sample.

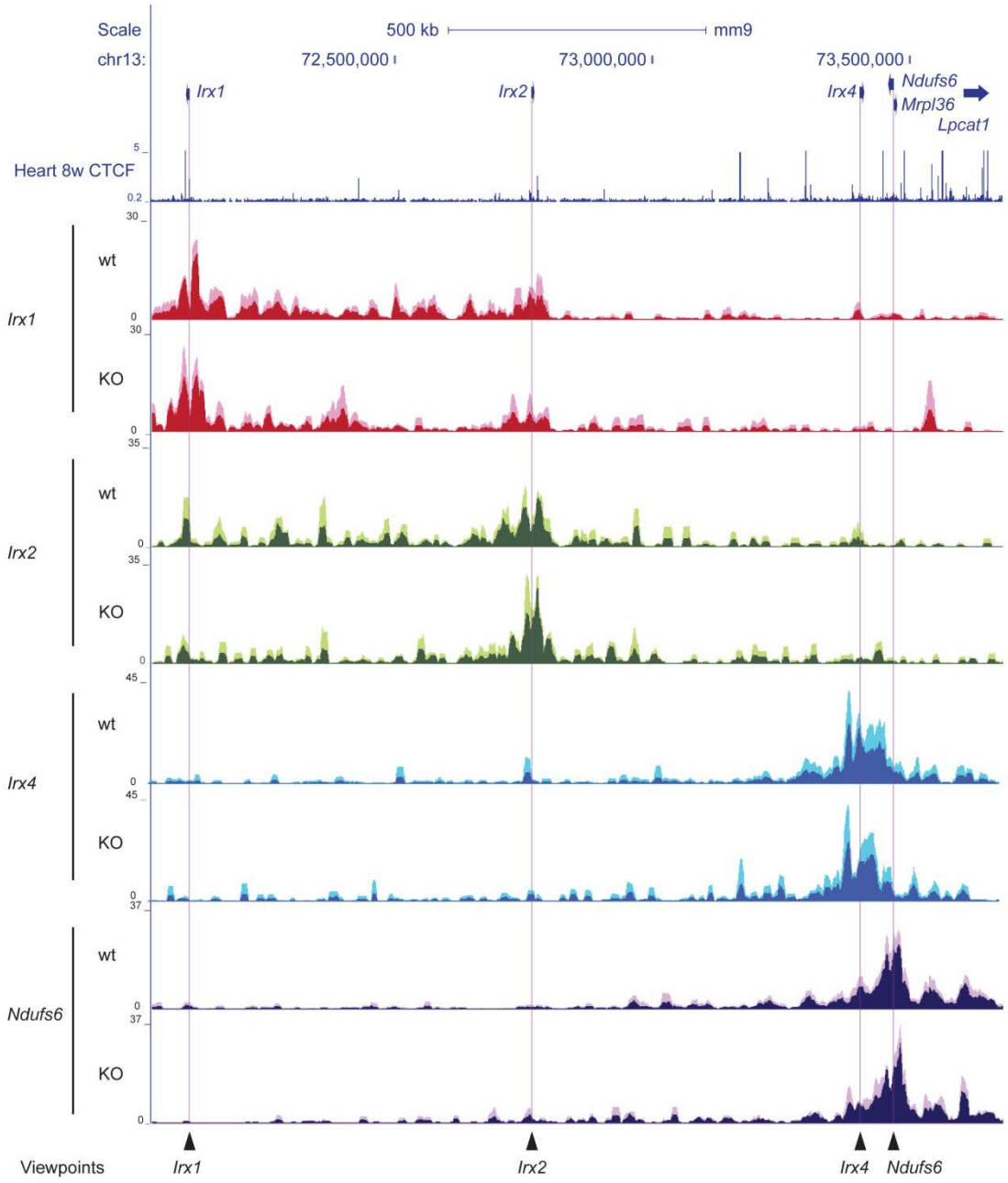

Fig. 33. Effect of *Ctcf* deletion on the 3D genomic structure of the IrxA cluster.

Chromatin interactions in the IrxA cluster in E11.5 control (wt) and KO (KO) hearts revealed by 4C-seq using viewpoints in the promoters of *Irx1*, *Irx2*, *Irx4* and *Ndufs6* (vertical purple lines). The binding profile of CTCF in 8-week-old hearts is shown at the top for reference (mm9; chr13:72,027,675-73,679,627). n≥3 for each sample.

We detected changes in interactions established by the promoter of *Irx4* with four specific genomic regions, two of them showing increased interaction in the Ctcf KO and two showing decreased interaction. The regions with increased interactions in the *Ctcf* KO contained a CTCF binding site (*i*) or, in the second case (*ii*), a previously described neural enhancer that has been shown to interact with the promoters of *Irx1* and *Irx2* but

not with the promoter of *Irx4* (Tena et al. 2011). The regions showing a decrease in the interactions with the *Irx4* promoter in the *Ctcf* KO surround the gene and contain CTCF binding sites. The region centromeric to the *Irx4* promoter (*iii*) harbored a candidate CTCF site that could be involved in separating the interaction and regulatory domain of the IrxA cluster. Indeed, we have shown that this region has insulator activity and prevents communication between regulatory elements in an *in vivo* zebrafish functional assay (appendix E). The second region with decreased interactions (*iv*) harbored several CTCF sites located at the 3' end of the *Lpcat1* gene. Although not statistical significant, two other interesting sites that showed decreased interaction with the *Irx4* promoter were the two downstream CTCF binding sites (* and **). The reason this sites are interesting is because they could be the limits of the *Irx4* regulatory landscape. In summary, the chromatin structure of the IrxA cluster shows clear alterations upon *Ctcf* deletion, and provides valuable insight to define the potential regulatory domain specific for *Irx4* in the heart.

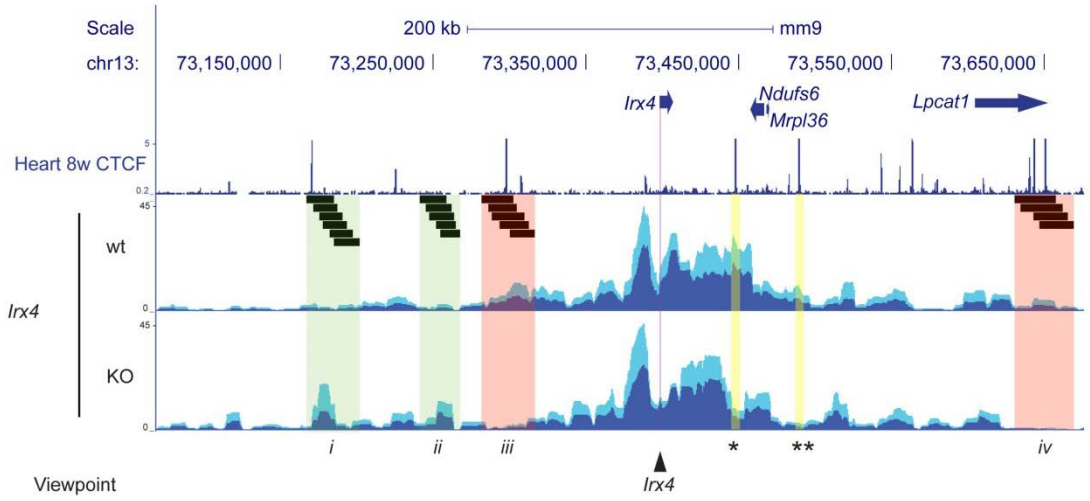

Fig. 34. Changes in the genomic structure of the *Irx4* locus upon *Ctcf* deletion.

4C-seq profiles using the *Irx4* promoter viewpoint (vertical purple line) in control (wt) and *Ctcf* KO (KO) embryonic hearts. Bins showing significant differences are indicated by short black lines and are boxed together, with shading in green when the interaction is gained in *Ctcf* KO (*i, ii*) and in red when it is lost (*iii, iv*). Yellow shading (*, **) highlights regions encompassing CTCF binding sites that show diminished interaction with the *Irx4* promoter in *Ctcf* KO, but that do not reach statistical significance (mm9: chr13:73,068,893-73,676,056). n=5 and n=3 for the control and *Ctcf* KO, respectively.

8. CTCF delimits the cardiac regulatory domain of *Irx4*

The detailed analysis of the interaction map for *Irx4* allowed us to delineate its putative regulatory landscape between the CTCF binding sites that showed reduced interaction in the *Ctcf* KO (Fig. 34). A strong prediction that follows is that regulatory elements

driving *Irx4* expression in the developing heart should be contained within this region. There are several well-characterized histone marks that help to locate active and inactive regions of the chromatin and identify those with potential enhancer activity (Shen et al. 2012).

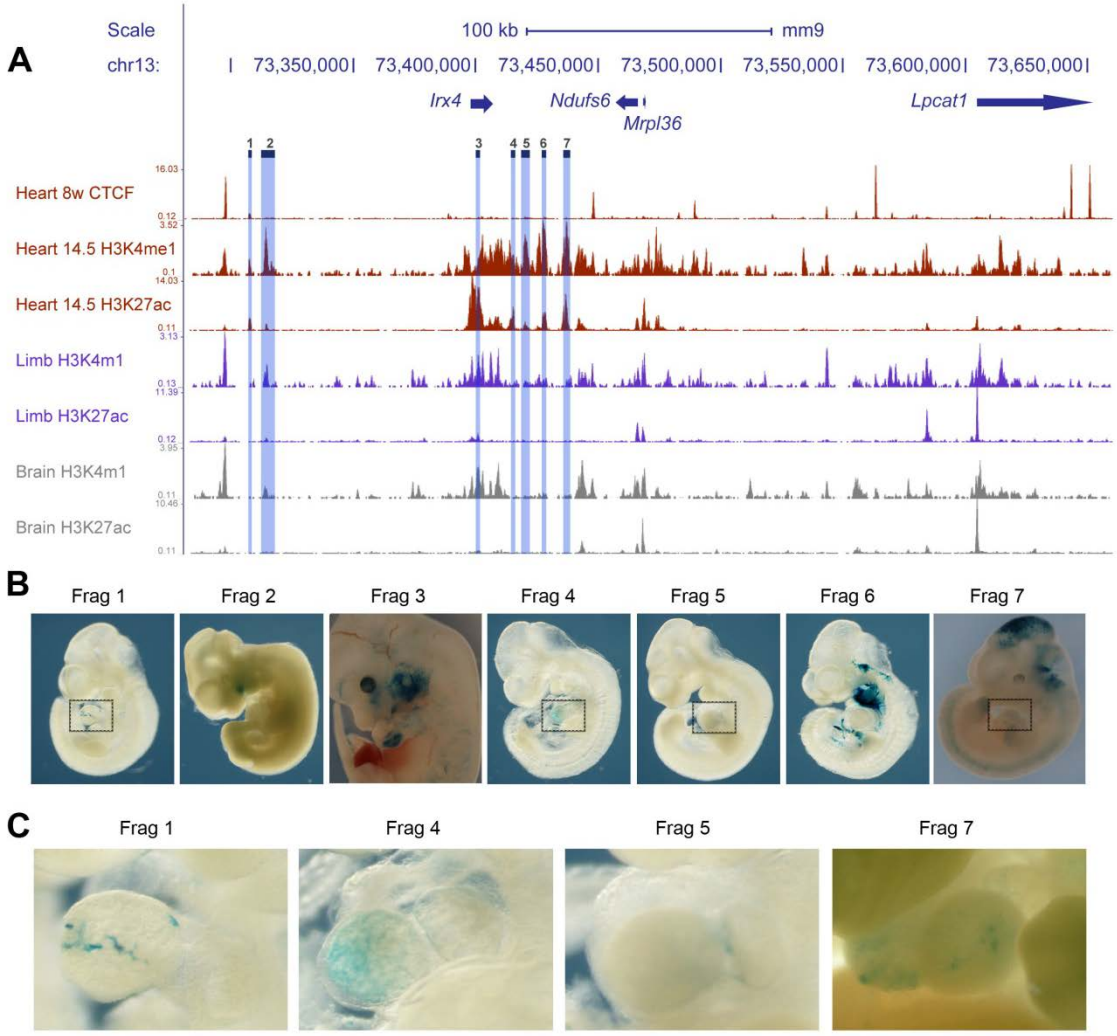

Fig. 35. *Irx4* putative regulatory landscape.

A, UCSC genome browser view of the *Irx4* region (mm9; chr13:73,281,240-73,660,195) showing (from top to bottom) the distribution of CTCF and histone modifications for active enhancers (H3K4me1 and H3K27ac) in the embryonic heart, limb and brain. Fragments selected for transgenic assays are highlighted in blue. **B,** LacZ reporter activity driven by the selected fragments (Frag 1-Frag 7) in transgenic embryos at E10.5 (Frag 1, 2, 4-7) or E11.5 (Frag 3). **C,** Higher magnification views of the boxed regions in **B**, showing reporter activity in the heart.

H3Kme1 marks active and poised enhancers while H327ac marks active elements (Rada-Iglesias et al. 2011). Within our putative *Irx4* regulatory landscape, we examined these two histone profiles in three mouse embryonic tissues: heart, brain and limb, as these tissues are expression territories of IrxA cluster genes during development

(Bosse et al. 1997). Based on the distribution of histone marks and excluding promoter regions, we chose seven candidate fragments to test their regulatory potential in a transient transgenic assay (Fig. 35A). All selected fragments presented H3K4me1 and H3K27ac histone marks in the heart. Only fragments 1, 4 and 5 exclusively presented this histone marks combination in the heart. Fragments 2 and 7 also presented this combination in the limb and H3K4me1 in the brain. Finally fragments 3 and 6, also presented H3K4me1 in the limb. No fragment presented H3K427ac in the brain (Table 15). We found that all seven fragments tested drove β-galactosidase expression in the embryo, demonstrating that these fragments have enhancer activity (Fig. 35B). However, only four of the regulatory elements (57%) induced β-galactosidase expression in the heart (Fig. 35C). These results show that there are several elements with heart enhancer potential inside the *Irx4* genomic region delimited by CTCF binding sites. Although all fragments drove expression in the embryo, only fragment 1 behaved as expected based on the histones profile. It presented the histone marks combination exclusively in the heart and we only detected enhancer activity in this organ. The rest of the fragments drove the reporter gene expression at least in one expected territory (Table 15).

9. *Ctcf* is necessary for the proper 3D assembly and functioning of the *Irx4* regulatory landscape

RNAseq and *in situ* hybridization showed that *Irx4* expression was downregulated in *Ctcf* KO hearts. Additionally, the interaction maps of *Irx1*, *Irx2* and *Irx4* promoters obtained by 4C-seq showed that the breakpoint of the interacting profiles occurred near the first centromeric CTCF binding site from *Irx4* (Fig. 34 *iii*). This site showed a decrease in interaction with the *Irx4* promoter in *Ctcf* KO hearts, and also showed insulator activity (appendix E). Taken together, these results indicate that this CTCF binding site (Fig. 34 *iii*) is a strong candidate that could demarcate the two interacting regions inside the IrxA cluster. Besides, when *Ctcf* is deleted, we observed a decrease in interaction between this site and the *Irx4* promoter, and we also observed *Irx4* downregulation. Thus, this site could be implicated in the 3D assembly of the *Irx4* landscape and in regulating its expression.

To gain further insight about this site in the structure of the locus, we performed 4C-seq using this site (Fig. 34 *iii*) as a viewpoint in control and *Ctcf* KO hearts. Although no significant differences were found, we detected regions showing weaker interactions, one with the *Irx4* promoter itself (*; Fig. 36), and with further CTCF sites downstream of *Irx4* (**, ***; Fig. 36). We also tested the interactions established from fragment 1

(frag1), that shows enhancer activity in the heart (Fig. 35B,C). We obtained similar results: reduced interaction with the *Irx4* promoter (*; Fig. 36), and with only one of the downstream CTCF binding sites (**; Fig. 36). Thus, *Ctcf* is necessary to form the interaction between this CTCF binding site and the *Irx4* promoter, and this interaction could be necessary to properly express *Irx4* in the developing heart.

Table 15. Correlation of histone marks and enhancer activity in transient transgenic embryos

frag	H3K4me1			H3K27ac			expression			
	heart	limb	brain	heart	limb	brain	heart	limb	brain	others
1	+			+			+			
2	+	+	+	+	+			+	+	+
3	+	+		+				+	+	+
4	+			+			+			+
5	+			+			+			+
6	+	+		+						+
7	+	+	+	+	+		+	+	+	

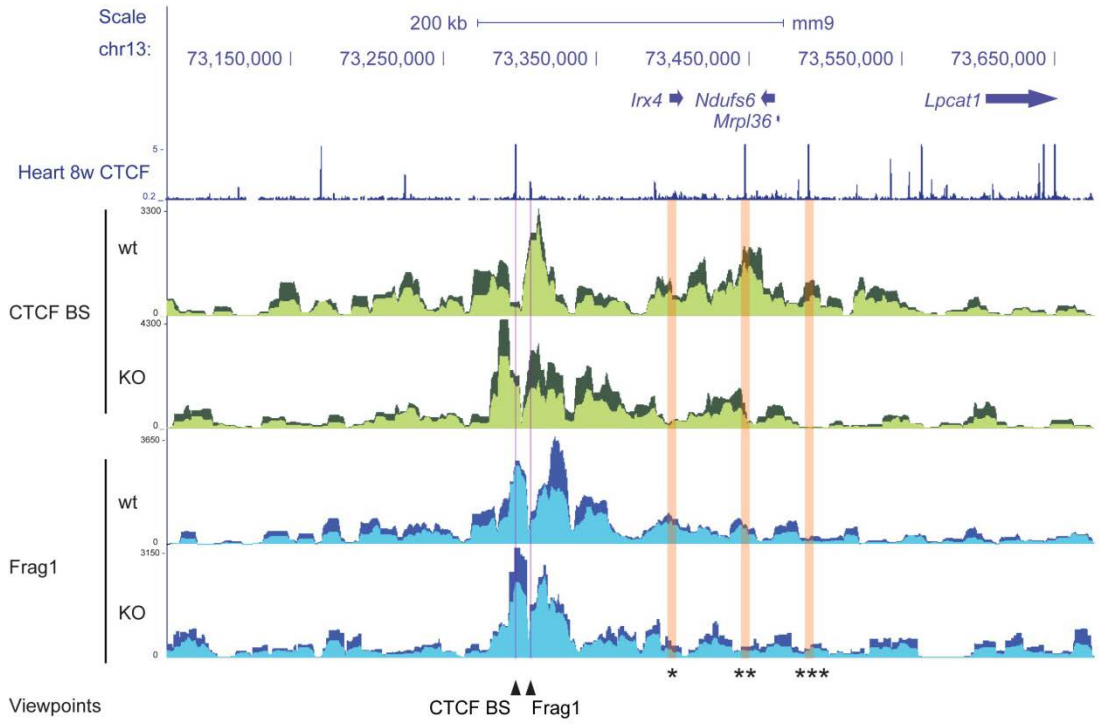

Fig. 36. *Ctcf* deletion impacts the 3D structure of the *Irx4* locus.

4C-seq in control (wt) and KO (KO) hearts using the CTFC binding site located between *Irx2* and *Irx4*, and the Frag1 enhancer as viewpoints (purple lines). Both viewpoints showed a decrease in interaction with the *Irx4* promoter (*) and the first (**) *Ctcf* binding site downstream *Irx4*. Only CTCF BS showed a decrease in interaction with a second CTCF binding site (***) The binding profile of CTCF (Shen et al. 2012) is shown at the top for reference (mm9; chr13:73,068,893-73,676,056). CTCF BS = CTCF binding site; Frag1 = fragment 1 n=2 for each sample.

The above results strongly suggest that the CTCF binding site between *Irx2* and *Irx4* (Fig. 34 *iii*) is involved in the 3D structure of the *Irx4* loci (Fig. 36). To determine if this site was also required for proper *Irx4* expression, we used the CRISPRs/Cas9 genome editing system to delete this CTCF binding site in a mouse transient transgenic assay (Fig. 37).

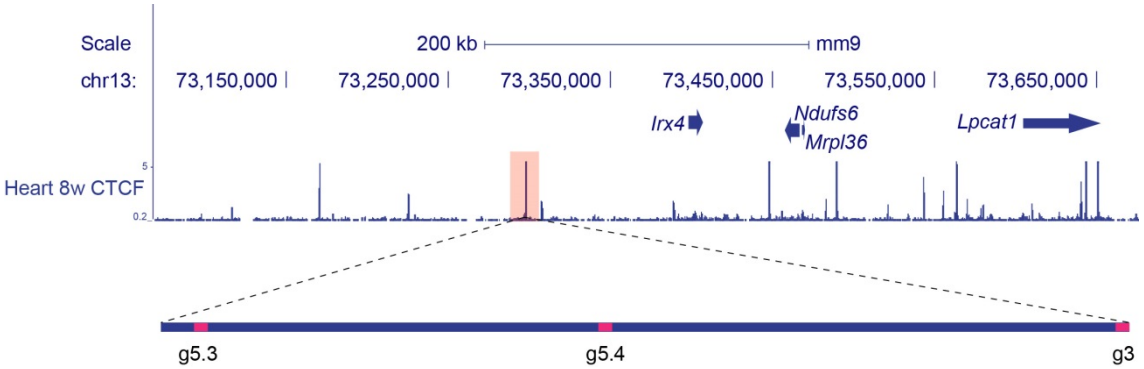

Fig. 37. Location of gRNAs used for the deletion of the *Irx2/Irx4* intergenic CTCF binding site.

Locations of the three guideRNAs used to delete the CTCF binding site. Two guide combinations were used: g5.3 and g3, and g5.4 and g3.

We designed three gRNA (guides RNA, see methods for details) that were used in two combinations (g5.3 and g3; g5.4 and g3, Fig. 38) to target the CTCF binding site (Fig. 34 *iii)*. One combination gave rise to a 1.3 kb deletion and the other combination gave rise to a smaller deletion of 779 bp (g5.3 and g3; g5.4 and g3, respectively). Each combination of gRNAs together with Cas9 protein was microinjected into the pronucleus of E0.5 mouse embryos. After transfer to foster CD1 mothers, they were collected at E9.5 or E10.5, processed for *in situ* RNA hybridization and their yolk sacs used for genotyping. Our genotyping strategy identified the presence of deleted alleles by PCR product size. The deleted allele from those embryos obtained with g5.3 and g3 combination had a PCR product of 829 bp, and the deleted allele from those embryos obtained with the g5.4 and g3 combination had a PCR product of 234 bp. Except for one embryo, we also detected the wildtype allele in all embryos genotyped, surely due to the expected mosaicism of the deletion at early stages. The sequencing confirmed the expected deletion and also showed that it was highly consistent (Fig. 38).

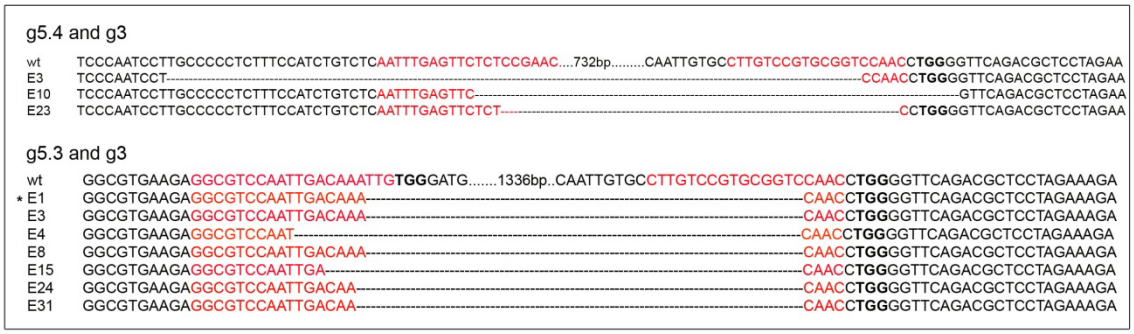

Fig. 38. CTFC binding site deletion using the CRISPR/Cas9 system.

A, Genomic sequence from embryos carrying the deletion generated using the combination of g5.4 and g3 gRNAs (779 bp deletion). **B,** Genomic sequence from embryos carrying the deletion generated using the combination of g5.3 and g3 gRNAs (1374 kb deletion).

Finally, to evaluate whether *Irx4* expression was altered when this CTCF binding site (Fig. 34*iii*) was deleted, we performed whole mount *Irx4 in situ* hybridization in embryos that were PCR-positive for the deletion, using CD1 embryos controls (referred here to as wild-type). We also performed in parallel *Irx4 in situ* hybridization for *Ctcf* KO embryos and its appropriate control in order to have a reference for *Irx4* downregulation upon *Ctcf* deletion. We found that *Irx4* was downregulated in the embryos that had a deletion of the CTCF binding site (Fig. 39A-D). This downregulation was more conspicuous than in *Cfcf* KO embryos (Fig. 39E-H). These results indicate that proper *Irx4* expression requires this *Ctcf* binding site (Fig. 34 *iii*). Collectively, we conclude that during heart development, the *Irx4* locus presents a CTCF-mediated 3D structure, which is required for the proper expression of the IrxA cluster and neighboring genes.

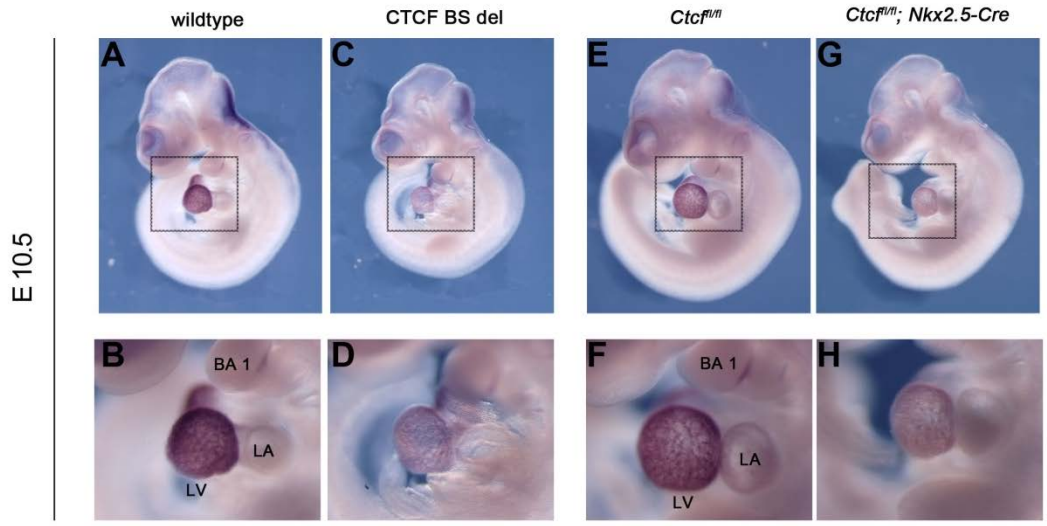

Fig. 39. The *Irx2/Irx4* intergenic CTCF binding site is necessary for proper *Irx4* expression.

A-D, *Irx4* whole mount *in situ* hybridization of E10.5 embryos where the *Irx2*/*Irx4* intergenic CTFC binding site has been deleted with the CRISPR/Cas9 system **A**, is a CD1 embryo (wild type). **C**, is the genome-edited embryo (CTCF BS del). **B** and **D** show close-up views of the hearts where *Irx4* downregulation upon CTCF site deletion is better appreciated. **E-H,** *Irx4* whole mount *in situ* hybridization of E10.5 embryos where the *Ctcf* gene was deleted using the *Cre*/LoxP system. **E**, is a control (*Ctcf*$^{fl/fl}$) embryo. **G**, is a *Ctcf* KO (*Ctcffl/fl;Nkx2.5-Cre*) embryo. **F** and **H** are close-up views of **E** and **G** respectively

DISCUSSION

Our understanding of the importance of 3D genomic structure for gene regulation is rapidly increasing through advances in new technologies (Bonora et al. 2014). The role of 3D structure during *in vivo* heart development, however, remains relatively unexplored (van Weerd et al. 2014). Since *Ctcf* is considered a master genome organizer in mammals (Phillips and Corces 2009) and increasing evidence supports the relevance of this protein in 3D structure (Ong and Corces 2014), we investigated the role of 3D structure for gene regulation during heart development by deleting *Ctcf* in developing cardiac tissue.

1. *Ctcf* is necessary for proper heart development

Because the germline *Ctcf* full knockout is lethal (Fedoriw et al. 2004; Moore et al. 2012; Wan et al. 2008), it must be conditionally deleted (Heath et al. 2008). We show here that the conditional deletion of *Ctcf* in the early cardiac lineage is embryonic-lethal, which is in accord with observations in other systems with tissue-specific deletion of *Ctcf* eg, in post-mitotic projection neurons (Hirayama et al. 2012), and early nervous system development (Watson et al. 2014). For example in B and T cells, proliferation, a key step in development, is impaired (Heath et al. 2008; Ribeiro de Almeida et al. 2011). Furthermore, in Th2 cells, the cytokine that confers Th2 cell identity is downregulated (Ribeiro de Almeida et al. 2009), and in myeloid cells, macrophage differentiation is impaired (Nikolic et al. 2014). In the early nervous system, deletion of *Ctcf* between E8 and E9 was found to reduce the size of the telencephalic and retinal structures, and no pups were recovered at birth. Moreover, shortly after the onset of deletion, at E11, an increase in apoptosis was detected. Deletion at later stages (E11) in the telencephalon produced a similar phenotype, but with a delay because the increase in apoptosis occurred at E14.5 and death was neonatal (Watson et al. 2014). *Ctcf* deletion in the oocyte delays the onset of mitosis and reduces meiosis competence. In addition, after fertilization, the second mitotic division is delayed, and zygote genome activation is perturbed. Embryos die before the blastocyst stage by apoptosis (Wan et al. 2008).

Because deletion of *Ctcf* in other developmental systems such as the oocyte, the limb and the nervous system, is known to increase apoptosis (Soshnikova et al. 2010; Wan et al. 2008; Watson et al. 2014) and leads to a decrease in cell proliferation in T cells (Heath et al. 2008; Ribeiro de Almeida et al. 2011), we analyzed these parameters in our conditional model. We chose to do so at the midpoint between the onset of deletion and embryonic death. In contrast to other systems, we failed to find any alterations in

these two processes. However, we do not rule out the possibility that apoptosis could be triggered later, perhaps at a time closer to embryonic death.

Our findings together with those of previous studies highlight the relevance of *Ctcf* in development; however, the final outcome of *Ctcf* deletion is not shared by all systems, as shown by the lack of effect in cellular processes such as cell proliferation or apoptosis ((Heath et al. 2008; Hirayama et al. 2012; Ribeiro de Almeida et al. 2011) and this work). Moreover, the phenotypes are different in each system studied, suggesting that the role of this protein is tissue specific. Indeed, it is interesting that the deletion of the same protein produces tissue-specific phenotypes, and that CTCF target genes such as Myc do not always respond to the absence of *Ctcf* (Wan et al. 2008). Although development is altered in most of the systems studied so far, it is unclear how tissue specificity is achieved.

Ctcf is known to associate with other proteins (Defossez et al. 2005; Donohoe et al. 2007), and posttranslational modifications regulate its binding to the chromatin (Klenova et al. 2001). Clearly, these events might be important for establishing tissue-specific roles for this protein. For example, the transcription factor YY1 can associate with CTCF and this associated complex plays a role in X chromosome inactivation (Donohoe et al. 2007). Interestingly, YY1 has a role in cardiac morphogenesis (Beketaev et al. 2015). However, deletion of this protein using the same *Nkx2*.5-*Cre* driver as used in this project altered cardiac cell proliferation and apoptosis (Beketaev et al. 2015), excluding it as a candidate to confer CTCF tissue-specific activity at least in the developing heart. In other systems, such as in B cells, these two proteins seem to interact in the same process of long-range interaction (Atchison 2014; Ribeiro de Almeida et al. 2011). Cohesin, another architectural protein, can bind to DNA alone or together with CTCF, mediator or Nipbl. In ES cells, the association of cohesin with Nipbl and mediator is related to the maintenance of the transcriptional program (Kagey et al. 2010). Thus, the search for CTCF cofactors and other posttranslational modifications that could influence its binding to DNA and/or its activity will help to better define its tissue specifificity.

2. *Ctcf* is necessary to activate the cardiac transcriptional program

To understand the role that *Ctcf* might play in heart development, we examined global transcriptional changes in our heart *Ctcf* KO model. GO term analysis of the comparison between *Ctcf* KO and control hearts showed that the GO terms related to

heart development and a group of genes related to developmental pathways such as Notch, TGFbeta, EGF/ErbB and BMP were present only in the downregulated genes in the *Ctcf* KO hearts. This suggests that in heart development, *Ctcf* is necessary to activate and/or maintain the cardiac transcriptional program. The upregulated genes included GO terms related to general metabolism. This dataset included ribosomal and mitochondrial proteins. Although no morphological defects were detected at the stage when we performed the RNAseq (E10.5), it seems that the downregulation of the cardiac transcriptional program is coincident with a metabolic change.

Other cases have been reported where *Ctcf* is necessary for activating transcription of genes required for development or function of the tissue studied (Delgado-Olguin et al. 2011; Heath et al. 2008; Hirayama et al. 2012; Ribeiro de Almeida et al. 2011). Among these, two published studies analyzed global transcriptional changes using microarray protocols. In one study, *Ctcf* was depleted in post-mitotic neurons (Hirayama et al. 2012), and in the other study *Ctcf* was knocked-down in zebrafish using morpholinos (Delgado-Olguin et al. 2011; Hirayama et al. 2012). In *Ctcf*-deficient post-mitotic neurons, genes related to cell adhesion, biological adhesion, behavior and post-embryonic development, all of which are associated with neural differentiation, were altered. The authors of this study focused on a group of genes, the protocadherins, because several members of this cluster are related to neural development and were downregulated (Hirayama et al. 2012). In the zebrafish model, several developmental GO terms were altered, and the authors focused on muscle development because it was in the top three GO terms detected. In contrast to our study, genes related to muscle development were up- and downregulated in *Ctcf* knocked-down zebrafish. This type of result also occurs in *Ctcf*-deleted macrophages, and genes related to macrophage function and development are up- and downregulated (Nikolic et al. 2014).

Collectively, these findings indicate that *Ctcf* regulates transcription and interestingly this occurs in a tissue-specific manner. The transcriptional changes in four systems, the heart (this work), limb, oocyte and cortex, are compared in Fig. 40. It should be noted that our DEG set was obtained from RNAseq analysis, whereas the DEG sets in the other systems were from microarrays, and consequently the total number of genes is different. Nonetheless, we would have expected that the DEG detected in the microarrays would be also found in our DEG from the RNAseq, however, this was not the case. The comparison shows that these systems share very few genes that are altered upon *Ctcf* deletion. This finding also strengthens the concept that CTCF displays tissue specificity. As discussed above, cofactors and posttranslational modifications could be involved in the tissue specificity of CTCF. The transcriptional

data suggest that this might involve transcription factors related to tissue identity, either by directly binding to CTCF or by activating the expression of cofactors or partners that assist CTCF to act in a tissue specific manner.

Additionally, in all cases studied transcription is altered globally. However, alteration of tissue-specific transcription programs as observed in zebrafish (Delgado-Olguin et al. 2011) and in this work, are also apparent. Moreover, more pronounced effects were found in specific loci functionally relevant to the tissue where *Ctcf* was deleted, as was the case of the protocadherins in post mitotic tissues (Hirayama et al. 2012), or the IL-10 locus in macrophages (Nikolic et al. 2014).

CTCF is an architectural and chromatin-bound protein with many thousands of binding sites scattered genome-wide (Ong and Corces 2014). The presence of a CTCF binding site in the vicinity of a gene suggests that it might play a role in regulating gene expression eg, as seen in the B-globin

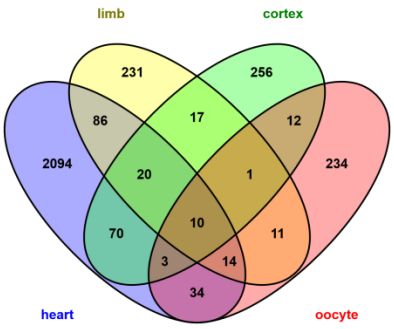

Fig. 40. Comparsion of the transcriptional changes upon *Ctcf* deletion in four systems.

We can see that the four systems are poorly overlapping.

locus (Splinter et al. 2012). In this work, we observed that in the *Ctcf* KO versus control comparison, DEG were closer to an embryonic CTCF binding site (E14.5) than genes that were expressed but unchanged, suggesting that *Ctcf* regulates their transcription. We also observed that the downregulated genes were closer to a predicted heart enhancer (Shen et al. 2012). Thus, given that *Ctcf* mediates long-range interactions (Ong and Corces 2014) that are necessary for bringing together enhancers and promoters in T cells (Shih et al. 2012), B cells (Ribeiro de Almeida et al. 2011), and other systems (Guo et al. 2012; Hirayama et al. 2012; Splinter et al. 2006), we propose that *Ctcf* activates the cardiac transcription program in the developing heart by bringing together heart enhancers to target genes.

3. *Ctcf* does not always act as an insulator in its genomic context

Insulators are DNA sequences that when placed between an enhancer and a promoter block the expression of the promoter (Wallace and Felsenfeld 2007). Several insulators have been described in vertebrates (Bell et al. 1999; Blackledge et al. 2007; Filippova et al. 2001; Giraldo et al. 2003; Majumder et al. 2006; Molto et al. 2009; Zhong and Krangel 1999). For the most part, insulator activity has been measured using in vitro methods, although some potential insulators have been tested in an in vivo zebrafish assay (Bessa et al. 2009).

It was appealing to hypothesize that *Ctcf* binding sites between genes with divergent expression in the developing heart would be good candidates to examine the role of *Ctcf* as an insulator. Indeed, this genomic configuration and expression pattern would suggest that these tandemly arranged genes (i.e., the troponins and Tbx gene pairs) do not share the same regulatory elements, and therefore *Ctcf* would function to prevent inadequate crosstalk between regulatory elements of one gene to act on its neighboring gene. This has been shown for the divergently regulated *Interleukin-3* and *Granulocyte macrophage colony-stimulating factor* genes (Bowers et al. 2009). In the case of the *Tbx3* and *Tbx5* gene pair, the interaction profile obtained by 4C-seq analysis was previously reported and indicated that indeed these two genes do not share regulatory elements, and that each one belongs to a different regulatory domain (van Weerd et al. 2014). Unexpectedly, our results showed that *Ctcf* binding sites located between genes with divergent expression patterns did not necessarily act as insulators since changes in expression patterns were not detected, only up- or downregulation in the same territory. Nevertheless, this does not automatically mean that other examples (not examined here) do have *Ctcf* insulator capacity.

The same logic was used in a previous study that attributed an insulator role for *Ctcf* when placed between genes with divergent expression; however, the authors did not test any of their insulators, *in vitro* or *in vivo* (Xie et al. 2007). There are several examples where CTCF binding site inactivation by methylation leads to aberrant gene expression, demonstrating that those sites can act as insulators (Filippova et al. 2001; Flavahan et al. 2016), and others where, based on a divergent expression pattern, we could expect that the absence of *Ctcf* or disruption of its binding site would lead to aberrant expression of genes separated by CTCF binding sites. For example, the HoxD cluster genes have a very distinctive spatial and temporal expression pattern. These genes are located in tandem and are characteristic because they are not all expressed at the same time. Rather, they are expressed sequentially according to their linear

arrangement in the genome. Moreover, their expression along the antero-posterior axis is also sequential and corresponds to their order in the genome. This is termed collinearity (Mark et al. 1997) and ChIP-seq analysis in limbs showed that CTCF binding sites were present between these genes. Assuming the role of CTCF as an insulator, we could expect that its removal would lead to mis-expression of these genes. However, conditional deletion of *Ctcf* in the limb failed to change the expression pattern of any of these genes (Soshnikova et al. 2010). In the case of the beta-globin locus, a CTCF binding site, 3´HS1, showed enhancer blocking activity in a transgene assay, but when the binding site was disrupted in vivo the expression of the neighboring genes was unaltered (Splinter et al. 2006).

These results suggest that the insulator activity mediated by CTCF is not the sole mechanism that restricts divergent expression patterns.

There is evidence that DNA sequences that do not bind CTCF can act as insulators. The locus control region (LCR) in the TCR locus has several hypersensitive sites (HS). One of these HS binds CTCF, but several others do not. Moreover, two HS that do not bind CTCF (HS4 and HS6) in the TCR locus insulate three co-expressed genes, TCRα, TCRδ and *Dad1* (Gomos-Klein et al. 2007; Magdinier et al. 2004). These genes are expressed in the same tissue, but at different stages during development. A similar case occurs in the *Rxbr* and *Col11a2* genes; DNA sequences are found between these two genes that show insulator activity but do not bind CTCF (Murai et al. 2008). An Alu element in the *K18* gene is another example (Willoughby et al. 2000). Other factors, such as SINE/ LINE, LTR and MIR elements, which are retrotransposable elements, could be responsible for isolating regulatory domains (Molto et al. 2009; J. Wang et al. 2015). In this work, we only searched for CTCF binding sites. We cannot discard the presence of other insulator sequences between the genes that are responsible for the insulation of regulatory domains. Another possibility could be that the interaction itself between the genes and their enhancers is what indirectly separates the regulatory domains.

4. The 3D structure of a paradigm cluster

The *iroquois* genes comprise a group of genes with a very specific genomic organization (Gomez-Skarmeta and Modolell 2002). Current evidence indicates that this genomic organization is relevant for their regulation (de la Calle-Mustienes et al. 2005; Tena et al. 2011). There are six *iroquois* genes distributed in two clusters, and each cluster has three genes. We analyzed the 3D structure of the cluster A.

Previous work showed that predominantly the first two members of the vertebrate *Iroquois* cluster A shared expression territories, and that *Irx4* had a more divergent and restricted expression pattern (Christoffels et al. 2000; Houweling et al. 2001). This arrangement is also present in other vertebrates, including *Xenopus*, Medaka (Tena et al. 2011) and even *Drosophila* (Cavodeassi et al. 2001). Moreover, enhancers tested so far from the IrxA cluster mainly interact with *Irx1* and *Irx2* (Tena et al. 2011). These findings strongly suggest that this cluster is divided into two regulatory domains (Tena et al. 2011). Our present results show that indeed the IrxA cluster is divided into two interacting domains of unequal size. The centromeric part of the cluster contains the *Irx1* and *Irx2* genes and the telomeric part contains the *Irx4* gene.

We also performed 4C-seq analysis of *Ndufs6*, the first gene downstream of the cluster A. *Ndufs6* is a subunit of mitochondrial Complex I. Considering their interaction profiles, *Irx4* and *Ndufs6* share interacting regions and must share some of the cardiac regulatory elements since they share expression territories as shown by RNA *in situ* hybridization. Based on the expression pattern and the genomic location of *Mrlp36*, we predict that its interaction profile will be highly overlapping at least with that of *Ndufs6*.

The IrxA cluster is an example of how expression patterns together with gene location can give insight into genomic organization. Shared expression patterns can predict, at least in part, the genomic regions where the regulatory elements acting on genes will be. This was also observed for the *Tbx3* and *Tbx5* genes; the fact that they lie next to each other and do not share expression patterns indicated different interacting domains and actually that was the case (van Weerd et al. 2014). The same interaction profiles can delimit the genomic region in which to look for regulatory elements. Consequently, genomic structure can provide information about the regulation of a gene. This delimitation in interaction between two genomic regions is a phenomenon that occurs genome-wide, and the IrxA cluster is an example of a TAD or two sub-TADs. The median size of a TAD is 880 kb (Dixon et al. 2012); however, they are also defined by the reduced interaction between two genomic regions (Valton and Dekker 2016). The IrxA cluster very likely organizes in a single TAD since there are regulatory elements shared by the three members of the cluster, and the two interacting regions are sub-TADs where enhancer-promoter interactions occur (Fig. 41).

5. The role of *Ctcf* in the 3D structure of the IrxA cluster

CTCF is an architectural protein (Ong and Corces 2014) that can bring together enhancers and promoters (Ribeiro de Almeida et al. 2011; Shih et al. 2012; Splinter et

al. 2006) and delimitate TADs (Dixon et al. 2012). Here, we show that the Irx cluster A has two interacting domains and the breakpoint of these domains has CTCF binding sites. Moreover, it was previously suggested that CTCF was implicated in the 3D structure of the loop formed by *Irx1/Irx2* (Tena et al. 2011). Interestingly, upon *Ctcf* deletion, only the *Irx4* gene presented a change in the 3D structure. The effect of this disruption was deregulation of genes from the cluster; *Irx4* was downregulated and *Irx1* was upregulated. The neighboring genes telomeric from the cluster, *Ndufs6*, *Mrlp36* and *Lpcat1*, were also upregulated. Furthermore, the new interaction formed between the *Irx4* promoter and a previously described neural enhancer that does not normally interact with this gene confirmed the disruption of the 3D structure of the cluster.

Our results show that genomic 3D structure in the developing heart plays a role in gene regulation as its disruption impacts this process. Nevertheless, our results with a *Ctcf* KO model suggest that the general structure appears to be very stable and is not easily disrupted. Two possibilities can be put forward to explain this.

The first possibility is that CTCF is not the only architectural protein involved in maintaining the 3D structure of the IrxA cluster or the genome in general, and in this setting other architectural proteins, such as cohesin, might be playing a role. Indeed, it has been previously shown that these two proteins can associate to promote long-range interaction and hence 3D structure (Degner et al. 2011; Guo et al. 2012; Merkenschlager and Odom 2013; Nasmyth and Haering 2009; Zuin et al. 2014). Moreover, cohesin has been shown to alter 3D structure and gene expression at specific loci including Tcra, the IFNG locus, protocadherins and the reg locus (Cuadrado et al. 2015; Hadjur et al. 2009; Seitan et al. 2011). In the case of the Tcra locus, the 3D structure was both CTCF and cohesin dependent (Hadjur et al. 2009). Although CTCF and cohesin are most commonly associated with long-range interaction, additional architectural proteins such as mediator (Kagey et al. 2010; Phillips-Cremins et al. 2013) and YY1 (Atchison 2014), among others, might act more locally but still have a role in 3D structure.

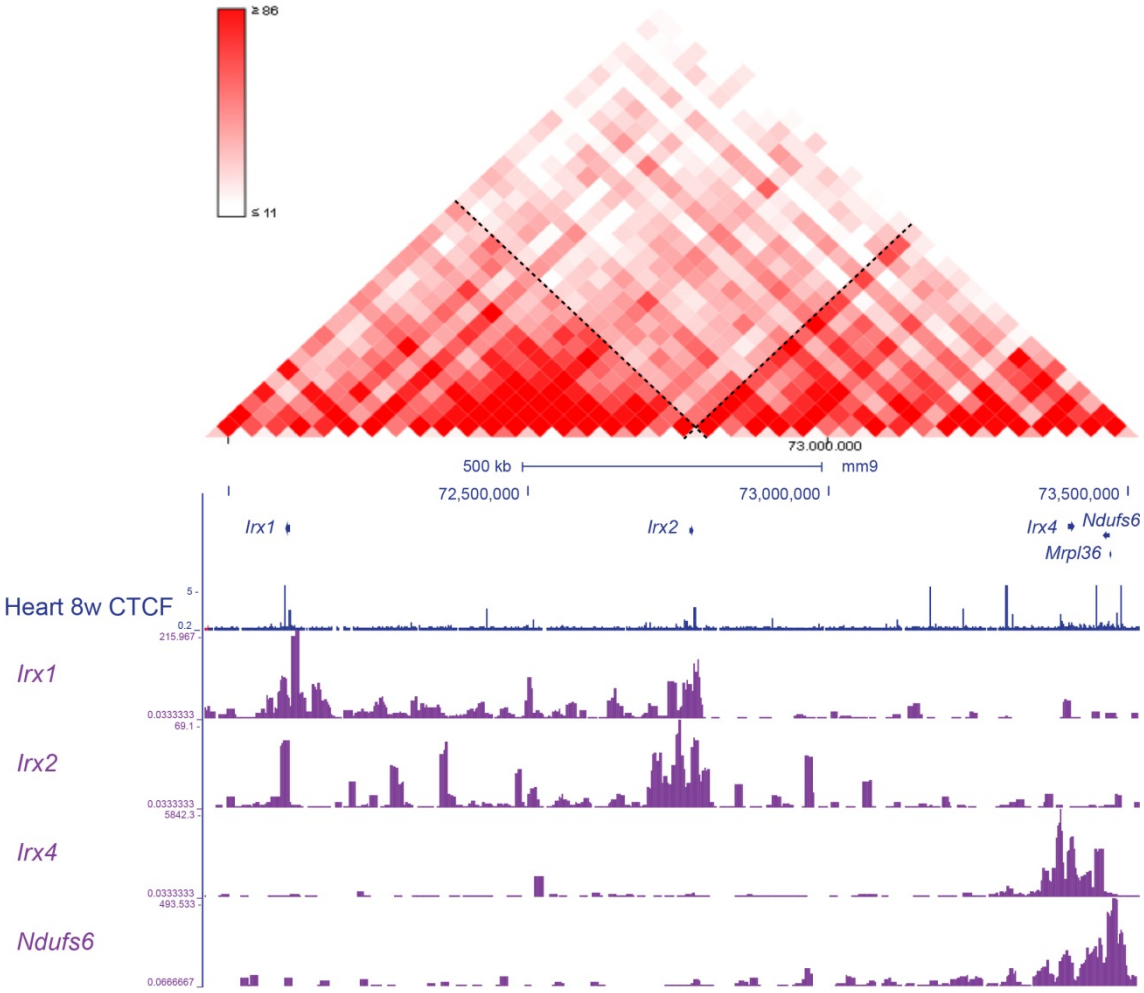

Fig. 41. The 3D structure of the IrxA cluster.

The IrxA cluster is divided into two subTADs. On top is a graphic representation of a Hi-C data from mouse ES cells (Dixon et al. 2012) where we can appreciate that the IrxA TAD seems to be divided in two subTADs. On the bottom the interaction profiles of the IrxA cluster genes and the first downstream gene from the Cluster A, *Ndufs6,* also show the subdivision of the IrxA cluster. The centromeric subTAD contains *Irx1* and *Irx2*. The telomeric subTAD contains *Irx4.*

Two reports suggest that the general 3D structure is very stable. Depletion of cohesin in non-cycling CD4 cells leads to gene expression alteration but the global 3D structure is maintained (Seitan et al. 2013). As in our work, the impact of the loss of another architectural protein, cohesin, seems to be local rather than global. A second study in HEK293T cells showed that the absence of CTCF or cohesin did not disrupt TADs (Zuin et al. 2014).

The second possibility is that indeed CTCF is necessary for the formation of the 3D structure of chromatin, but dispensable for its later maintenance. Hence, structures that change dynamically during development could be more susceptible to CTCF depletion; such would be the case for the recombination locus in B cells (Ribeiro de Almeida et al. 2011), and promoter choice in the protocadherin alpha cluster (Hirayama et al. 2012). If

the second possibility is correct, we should expect changes in the 3D structure of cluster A in general. However, the *iroquois* genes have a dynamic expression pattern in other tissues including neural tissue and limbs. Conditional deletion of *Ctcf* in these contexts might answer this question. In our model, we deleted *Ctcf* in early heart development, before cardiac tissue specification but after mesoderm differentiation. Therefore, the chromatin very likely already has a 3D structure. Deletion of *Ctcf* at earlier stages in development could lead to embryonic lethality before cardiac specification is achieved. Thus, it is difficult to test if *Ctcf* is necessary for the formation of 3D structure in our working model.

6. CTCF binding sites delimit the *Irx4* regulatory landscape

Given that *Irx4* is in a different regulatory domain and is highly expressed in the heart, we hypothesized that its regulatory domain should harbor cardiac regulatory elements. Although the *Irx4* interaction profile illustrates its interacting and hence regulatory domain, the precise limits were not entirely clear. However, the interactions centromeric to the gene seemed to decrease near CTCF binding sites. Moreover, a neural enhancer that did not show interaction with *Irx4* under wild-type conditions (Tena et al. 2011) was upstream from the first CTCF binding site centromeric to *Irx4*. This site showed insulator activity in a zebrafish insulator assay (appendix E). Thus, we propose that this site is the upstream limit of the *Irx4* regulatory domain.

The downstream limit was less clear. The first decrease in interaction also occurred at a CTCF binding site telomeric to the gene and located proximal to *Ndufs6*. A different option for the *Irx4* downstream limit lies further away and could terminate near the *Lpcat1* promoter, for several reasons. First, the interaction map extends this far. Second, the promoter of *Lpcat1* is in a genomic region that loses interaction with *Irx4* upon *Ctcf* depletion. Finally, this site harbors several CTCF binding sites that could act as boundaries. Since the upstream boundary is disrupted upon *Ctcf* deletion, this could also be a possibility for the downstream boundary. Further, as shown by our RNAseq analysis, the *Irx4* downstream neighboring genes *Ndufs6*, *Mrlp36* and *Lpcat1* are expressed in the heart. Moreover, *in situ* hybridization showed that *Ndufs6* and *Mrlp36* share *Irx4* expression domains, and the 4C-seq profile of *Irx4* and *Ndufs6* demonstrated that they share interacting regions. Taken together, these findings suggest that all four genes can share regulatory elements in the 350 kb region where they are located. Thus, the *Lpcat1* promoter could be the *Irx4* downstream regulatory landscape limit.

Previous work used DNA conservation to search for regulatory elements inside the IrxA cluster; however, none of the elements tested drove expression in the heart (Tena et al. 2011), raising the possibility that although conservation is useful to find regulatory elements, it may have to be complemented with other strategies (Blow et al. 2010; Visel et al. 2009). We followed the previously described strategy that uses histone marks to search for regulatory elements since histone modifications are associated with promoters, enhancers and active regions of DNA (H3K4me3, H3K4me1 and H3K27ac, respectively (Bernstein et al. 2006; Creyghton et al. 2010; Rada-Iglesias et al. 2011)). We aimed to uncover *Irx4* regulatory elements and we searched for them in the 150 kb region flanked by CTCF binding sites. But, none of the enhancers tested recapitulated the whole *Irx4* expression pattern. Nevertheless, 57% of the elements tested drove expression of the *LacZ* reporter gene in the heart. Thus, cardiac regulatory elements are indeed present in the region analyzed. We do not discard the possibility that enhancers that more precisely reproduce *Irx4* expression exist in this 150 kb region and that our strategy failed to find these elements. Clearly, alternative explanations for this lack of equivalency could be that those regulatory elements are located further away from the limited region we established, or that the combination of the enhancers identified is necessary to reproduce the full *Irx4* expression pattern,.

7. Looping *Irx4*

Previous studies have shown that interaction between CTCF binding sites and promoters is necessary for gene expression (Ribeiro de Almeida et al. 2011; Shih et al. 2012). Our results suggest that the *Irx4* promoter interacts with CTCF binding sites upstream and downstream of the gene and that these interactions, besides delimiting the regulatory landscape, could be related to *Irx4* regulation because when *Irx4* is downregulated, there is also a decrease in interaction between the promoter and these CTCF binding sites. Our results suggest that the CTCF upstream site was a boundary for the *Irx4* regulatory landscape. The interacting profile using this CTCF binding site as the viewpoint showed that it contacted the *Irx4* promoter and other CTCF binding sites located downstream of *Irx4*. Although not statistically significant, these interactions were lost upon *Ctcf* deletion. Thus, this binding site interacts with the *Irx4* promoter and very likely this interaction regulates *Irx4* gene expression. The interacting profile from a close-by enhancer showed that it contacted the *Irx4* promoter and that this interaction was also lost upon *Ctcf* deletion. Collectively, these findings strongly indicate that the 3D structure of the *Irx4* locus is CTCF dependent.

Finally, we deleted the CTCF binding site that we defined as the upstream limit of *Irx4* regulatory landscape. We did this using the CRIPSR/Cas9 system in transient transgenic assays. Interestingly, the deletion of this binding site downregulated *Irx4* expression only in the heart. This result has several implications. First, this CTCF binding site is directly associated with *Irx4* expression and positively regulates this gene. This site loops towards the *Irx4* promoter and in doing so likely brings regulatory elements to the gene. If this is the case, it provides an explanation for the downregulation of *Irx4* when *Ctcf* or the CTCF-binding site is deleted. This site is located near the point that separates the two interacting regions of cluster A and is therefore an attractive candidate for a boundary element inside the cluster. Moreover, it could also be acting as an insulator since it separates genes with divergent expression patterns. If this site represented an actual insulator, we would expect a "contamination" of the regulatory domains inside the cluster A, and thus ectopic expression of *Irx4*. This ectopic expression would be expected in the expression territories of *Irx1* and *Irx2*, which include the nervous system and limbs. However, *Irx4* was not ectopically expressed in any edited embryo analyzed, indicating that it does not function as an insulator. Although the transient transgenic embryos were mosaic, they did not have the CTCF binding deletion restricted to the heart or any other tissue. It is conceivable that the looping from the genomic region close to the location of this CTCF binding site towards both sides of the cluster is actually what separates the regulatory domains. For example, the neural enhancer upstream of this site interacts with *Irx1* and *Irx2* promoters (Tena et al. 2011) and the CTCF binding site (the one deleted) interacts with the *Irx4* promoter. Both loops are established to regulate gene expression; they bring together regulatory elements and target promoters and, at the same time, restrict interaction between *Irx1/Irx2* and *Irx4* promoters. Our data allow us to propose that the *Irx4* 3D structure is CTCF mediated, and that this genomic region forms at least two loops with the CTCF binding sites upstream and downstream from the gene. In doing so, this 3D structure regulates *Irx4* gene expression in the developing mammalian heart (Fig. 42).

Using the IrxA cluster as an example, we show that 3D genomic structure is implicated in gene regulation and that CTCF is one of the players mediating 3D genomic structure during heart development.

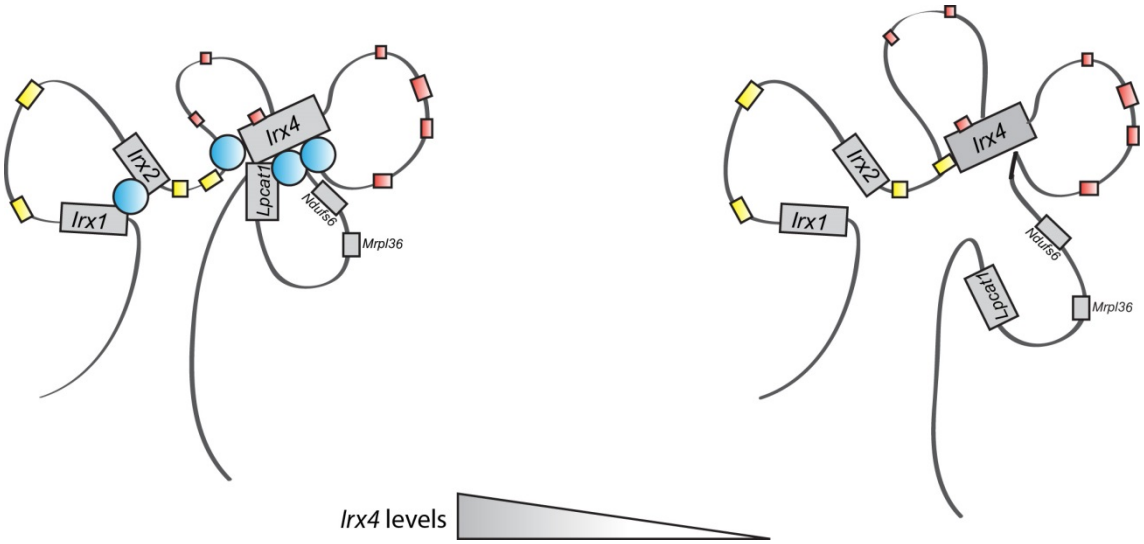

Fig. 42. CTCF mediates *Irx4* 3D structure.

The 3D structure of the IrxA cluster is depicted in the diagram on the left. Blueish circles represent CTCF and rectangles are the enhancers tested so far either in this work (red) or in previous works (yellow) (Tena et al. 2011). The diagram on the right depicts the 3D structure upon *Ctcf* deletion (represented by the absence of blueish circles). The loop containing *Irx1* and *Irx2* remains stable, whereas the 3D structure for proper *Irx4* expression is dismantled. *Irx4* loses interaction with the *Lpcat1* promoter and at the same time gains interaction with a neural enhancer (yellow rectangle next to *Irx4*).

In this work, we show that the architectural protein CTCF helps to activate the cardiac transcription program in the developing heart. It does so by facilitating enhancer-promoter interaction. Although CTCF can act as an insulator, it seems that other elements and mechanisms can separate regulatory domains, as in the case of *Tbx* and troponin genes. Our work also highlights that genes with divergent gene expression patterns very likely belong to different regulatory domains; such is the case for the cluster A *iroquois* genes. Moreover, we found that the global scaffolding of the chromatin is resilient to dismantling. However, there are regulatory landscapes susceptible to disassembly upon *Ctcf* loss. In conclusion the 3D structure is relevant and necessary for gene regulation in the developing mammalian heart. How CTCF reaches tissue specificity remains an open and intriguing question. Future work will help to find CTCF tissue specific partners and whether they help CTFC to stablish 3D structure for proper development and homeostasis maintenance.

CONCLUSIONS

1. *Ctcf* is necessary for heart development and its deletion in the mouse embryonic heart is embryonic lethal.

2. Deletion of *Ctcf* in the developing heart causes severe cardiac malformations that include shortening of the interventricular septum and reduced thickness of the compact myocardial wall.

3. Before cardiac malformations arise, deletion of *Ctcf* in the mouse embryonic heart does not alter cell proliferation or apoptosis.

4. *Ctcf* is necessary for the activation of the cardiac developmental transcriptional program and it possibly does so by bringing together heart enhancers and target genes.

5. The CTCF binding sites located between genes with divergent expression patterns do not always act as insulators in their genomic context.

6. The IrxA cluster is divided into two regulatory landscapes of unequal size. The centromeric regulatory domain includes *Irx1* and *Irx2*, and the telomeric domain includes *Irx4*.

7. Loss of *Ctcf* in the mouse embryonic heart disrupts the 3D structure of the IrxA gene Cluster, and this alters gene expression of the IrxA cluster and neighboring genes.

8. The regulatory landscape of *Irx4* is delimited by CTCF binding sites and this genomic region contains regulatory elements that drive expression into the mouse embryonic heart.

9. A specific CTCF binding site located between *Irx2* and *Irx4* is involved in establishing the proper 3D structure of the *Irx4* regulatory landscape and is necessary for its correct expression.

CONCLUSIONES

1. *Ctcf* es necesario para el desarrollo cardiaco y su deleción en el corazón embrionario de ratón causa letalidad embrionaria.

2. La deleción de *Ctcf* en el corazón en desarrollo causa malformaciones cardiacas severas, que incluyen un acortamiento del septo interventricular y una reducción en el grosor de la pared del miocardio compacto.

3. Antes de que se detecten las malformaciones cardiacas, la deleción de *Ctcf* en el corazón embrionario de ratón no altera la proliferación celular ni la apoptosis.

4. *Ctcf* es necesario para la activación del programa transcripcional de desarrollo cardiaco, y es posible que realice esta función al aproximar enhancers de corazón con sus genes blanco.

5. Los sitios de unión a CTCF localizados entre genes con patrones de expresión divergentes no siempre actúan como aisladores en su contexto genómico.

6. El cluster IrxA está dividido en dos regiones regulatorias de distinto tamaño. El dominio regulatorio centromérico incluye *Irx1* e *Irx2*, y el dominio telomérico incluye *Irx4*.

7. La pérdida de *Ctcf* en el corazón embrionario de ratón deshace la estructura 3D de los genes del cluster IrxA, y esto altera la expresión génica del cluster IrxA y de genes vecinos.

8. La región regulatoria de *Irx4* está delimitada por sitios de unión a CTCF, y esta región genómica contiene elementos reguladores que dirigen la expresión al corazón embrionario de ratón.

9. Un sitio específico de unión a CTCF localizado entre *Irx2* e *Irx4* está involucrado en establecer la estructura 3D apropiada de la región regulatoria de *Irx4,* y es necesario para que se exprese correctamente.

BIBLIOGRAPHY

Aguirre, L. A., et al. (2015), 'Long-range regulatory interactions at the 4q25 atrial fibrillation risk locus involve PITX2c and ENPEP', *BMC Biol,* 13, 26.

Alexander, J. M., et al. (2015), 'Brg1 modulates enhancer activation in mesoderm lineage commitment', *Development,* 142 (8), 1418-30.

Alexander, R. P., et al. (2010), 'Annotating non-coding regions of the genome', *Nature reviews. Genetics,* 11 (8), 559-71.

Anderson, E., et al. (2014), 'Mapping the Shh long-range regulatory domain', *Development,* 141 (20), 3934-43.

Atchison, M. L. (2014), 'Function of YY1 in Long-Distance DNA Interactions', *Frontiers in immunology,* 5, 45.

Bao, Z. Z., et al. (1999), 'Regulation of chamber-specific gene expression in the developing heart by Irx4', *Science,* 283 (5405), 1161-4.

Barton, P. J., et al. (2004), 'The slow skeletal muscle troponin T gene is expressed in developing and diseased human heart', *Molecular and cellular biochemistry,* 263 (1-2), 91-7.

Beketaev, I., et al. (2015), 'Critical role of YY1 in cardiac morphogenesis', *Developmental dynamics : an official publication of the American Association of Anatomists,* 244 (5), 669-80.

Bell, A. C., West, A. G., and Felsenfeld, G. (1999), 'The protein CTCF is required for the enhancer blocking activity of vertebrate insulators', *Cell,* 98 (3), 387-96.

Bernstein, B. E., et al. (2006), 'A bivalent chromatin structure marks key developmental genes in embryonic stem cells', *Cell,* 125 (2), 315-26.

Bessa, J., et al. (2009), 'Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish', *Developmental dynamics : an official publication of the American Association of Anatomists,* 238 (9), 2409-17.

Bickmore, W. A. (2013), 'The spatial organization of the human genome', *Annual review of genomics and human genetics,* 14, 67-84.

Blackledge, N. P., Rose, N. R., and Klose, R. J. (2015), 'Targeting Polycomb systems to regulate gene expression: modifications to a complex story', *Nature reviews. Molecular cell biology,* 16 (11), 643-9.

Blackledge, N. P., et al. (2007), 'CTCF mediates insulator function at the CFTR locus', *The Biochemical journal,* 408 (2), 267-75.

Blow, M. J., et al. (2010), 'ChIP-Seq identification of weakly conserved heart enhancers', *Nature genetics,* 42 (9), 806-10.

Bonora, G., Plath, K., and Denholtz, M. (2014), 'A mechanistic link between gene regulation and genome architecture in mammalian development', *Current opinion in genetics & development,* 27, 92-101.

Bosse, A., et al. (1997), 'Identification of the vertebrate Iroquois homeobox gene family with overlapping expression during early development of the nervous system', *Mechanisms of development,* 69 (1-2), 169-81.

Bowers, S. R., et al. (2009), 'A conserved insulator that recruits CTCF and cohesin exists between the closely related but divergently regulated interleukin-3 and granulocyte-macrophage colony-stimulating factor genes', *Molecular and cellular biology,* 29 (7), 1682-93.

Bruneau, B. G. (2002), 'Transcriptional regulation of vertebrate cardiac morphogenesis', *Circulation research,* 90 (5), 509-19.

Bruneau, B. G., et al. (1999), 'Chamber-specific cardiac expression of Tbx5 and heart defects in Holt-Oram syndrome', *Developmental biology,* 211 (1), 100-8.

Bruneau, B. G., et al. (2001), 'Cardiomyopathy in Irx4-deficient mice is preceded by abnormal ventricular gene expression', *Molecular and cellular biology,* 21 (5), 1730-6.

Buckingham, M., Meilhac, S., and Zaffran, S. (2005), 'Building the mammalian heart from two sources of myocardial cells', *Nature reviews. Genetics,* 6 (11), 826-35.

Cavodeassi, F., Modolell, J., and Gomez-Skarmeta, J. L. (2001), 'The Iroquois family of genes: from body building to neural patterning', *Development,* 128 (15), 2847-55.

Clowes, C., et al. (2014), 'The functional diversity of essential genes required for mammalian cardiac development', *Genesis,* 52 (8), 713-37.

Costantini, D. L., et al. (2005), 'The homeodomain transcription factor Irx5 establishes the mouse cardiac ventricular repolarization gradient', *Cell,* 123 (2), 347-58.

Crane, E., et al. (2015), 'Condensin-driven remodelling of X chromosome topology during dosage compensation', *Nature,* 523 (7559), 240-4.

Cremer, T. and Cremer, C. (2001), 'Chromosome territories, nuclear architecture and gene regulation in mammalian cells', *Nature reviews. Genetics,* 2 (4), 292-301.

Creyghton, M. P., et al. (2010), 'Histone H3K27ac separates active from poised enhancers and predicts developmental state', *Proceedings of the National Academy of Sciences of the United States of America,* 107 (50), 21931-6.

Cuadrado, A., et al. (2015), 'The contribution of cohesin-SA1 to gene expression and chromatin architecture in two murine tissues', *Nucleic acids research,* 43 (6), 3056-67.

Chao, W., et al. (2002), 'CTCF, a candidate trans-acting factor for X-inactivation choice', *Science,* 295 (5553), 345-7.

Chapman, D. L., et al. (1996), 'Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development', *Developmental dynamics : an official publication of the American Association of Anatomists,* 206 (4), 379-90.

Chen, H., et al. (2012a), 'Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome', *PloS one,* 7 (7), e41374.

Chen, L., et al. (2012b), 'Conditional ablation of Ezh2 in murine hearts reveals its essential roles in endocardial cushion formation, cardiomyocyte proliferation and survival', *PloS one,* 7 (2), e31005.

Christoffels, V. M., et al. (2000), 'Patterning the embryonic heart: identification of five mouse Iroquois homeobox genes in the developing heart', *Developmental biology,* 224 (2), 263-74.

Christoffels, V. M., et al. (2009), 'Tbx18 and the fate of epicardial progenitors', *Nature,* 458 (7240), E8-9; discussion E9-10.

Christoffels, V. M., et al. (2006), 'Formation of the venous pole of the heart from an Nkx2-5-negative precursor population requires Tbx18', *Circulation research,* 98 (12), 1555-63.

de la Calle-Mustienes, E., et al. (2005), 'A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts', *Genome Res,* 15 (8), 1061-72.

de Wit, E. and de Laat, W. (2012), 'A decade of 3C technologies: insights into nuclear organization', *Genes & development,* 26 (1), 11-24.

de Wit, E., et al. (2015), 'CTCF Binding Polarity Determines Chromatin Looping', *Molecular cell,* 60 (4), 676-84.

de Wit, E., et al. (2013), 'The pluripotent genome in three dimensions is shaped around pluripotency factors', *Nature,* 501 (7466), 227-31.

Defossez, P. A., et al. (2005), 'The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso', *The Journal of biological chemistry,* 280 (52), 43017-23.

Degner, S. C., et al. (2011), 'CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells', *Proceedings of the National Academy of Sciences of the United States of America,* 108 (23), 9566-71.

Dekker, J., et al. (2002), 'Capturing chromosome conformation', *Science,* 295 (5558), 1306-11.

Delgado-Olguin, P., et al. (2011), 'CTCF promotes muscle differentiation by modulating the activity of myogenic regulatory factors', *The Journal of biological chemistry,* 286 (14), 12483-94.

Dixon, J. R., et al. (2012), 'Topological domains in mammalian genomes identified by analysis of chromatin interactions', *Nature,* 485 (7398), 376-80.

Donohoe, M. E., et al. (2007), 'Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch', *Molecular cell,* 25 (1), 43-56.

Fedoriw, A. M., et al. (2004), 'Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting', *Science,* 303 (5655), 238-40.

Ferrai, C., et al. (2010), 'Gene positioning', *Cold Spring Harbor perspectives in biology,* 2 (6), a000588.

Filippova, G. N., et al. (2001), 'CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus', *Nature genetics,* 28 (4), 335-43.

Flavahan, W. A., et al. (2016), 'Insulator dysfunction and oncogene activation in IDH mutant gliomas', *Nature,* 529 (7584), 110-4.

Franco, D. and Campione, M. (2003), 'The role of Pitx2 during cardiac development. Linking left-right signaling and congenital heart diseases', *Trends in cardiovascular medicine,* 13 (4), 157-63.

Gaborit, N., et al. (2012), 'Cooperative and antagonistic roles for Irx3 and Irx5 in cardiac morphogenesis and postnatal physiology', *Development,* 139 (21), 4007-19.

Garcia-Gonzalez, E., et al. (2016), 'Chromatin remodeling effects on enhancer activity', *Cellular and molecular life sciences : CMLS*.

Ghavi-Helm, Y., et al. (2014), 'Enhancer loops appear stable during development and are associated with paused polymerase', *Nature,* 512 (7512), 96-100.

Giraldo, P., et al. (2003), 'Functional dissection of the mouse tyrosinase locus control region identifies a new putative boundary activity', *Nucleic acids research,* 31 (21), 6290-305.

Gomez-Marin, C., et al. (2015), 'Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders', *Proceedings of the National Academy of Sciences of the United States of America,* 112 (24), 7542-7.

Gomez-Skarmeta, J. L. and Modolell, J. (2002), 'Iroquois genes: genomic organization and function in vertebrate neural development', *Current opinion in genetics & development,* 12 (4), 403-8.

Gomos-Klein, J., et al. (2007), 'CTCF-independent, but not CTCF-dependent, elements significantly contribute to TCR-alpha locus control region activity', *Journal of immunology,* 179 (2), 1088-95.

Greulich, F., Rudat, C., and Kispert, A. (2011), 'Mechanisms of T-box gene function in the developing heart', *Cardiovascular research,* 91 (2), 212-22.

Guo, Y., et al. (2012), 'CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice', *Proceedings of the National Academy of Sciences of the United States of America,* 109 (51), 21081-6.

Guo, Y., et al. (2015), 'CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function', *Cell,* 162 (4), 900-10.

Hadjur, S., et al. (2009), 'Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus', *Nature,* 460 (7253), 410-3.

Handoko, L., et al. (2011), 'CTCF-mediated functional chromatin interactome in pluripotent cells', *Nature genetics,* 43 (7), 630-8.

Hark, A. T., et al. (2000), 'CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus', *Nature,* 405 (6785), 486-9.

Heath, H., et al. (2008), 'CTCF regulates cell cycle progression of alphabeta T cells in the thymus', *The EMBO journal,* 27 (21), 2839-50.

Hirayama, T., et al. (2012), 'CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons', *Cell reports,* 2 (2), 345-57.

Hoogaars, W. M., et al. (2004), 'The transcriptional repressor Tbx3 delineates the developing central conduction system of the heart', *Cardiovascular research,* 62 (3), 489-99.

Horton, A. C., et al. (2008), 'Conservation of linkage and evolution of developmental function within the Tbx2/3/4/5 subfamily of T-box genes: implications for the origin of vertebrate limbs', *Development genes and evolution,* 218 (11-12), 613-28.

Houweling, A. C., et al. (2001), 'Gene and cluster-specific expression of the Iroquois family members during mouse development', *Mechanisms of development,* 107 (1-2), 169-74.

Huang, Q. Q., et al. (2008), 'Co-expression of skeletal and cardiac troponin T decreases mouse cardiac function', *American journal of physiology. Cell physiology,* 294 (1), C213-22.

Jinek, M., et al. (2012), 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science,* 337 (6096), 816-21.

Jostarndt, K., et al. (1994), 'The use of 33P-labelled riboprobes for *in situ* hybridizations: localization of myosin alkali light-chain mRNAs in adult human skeletal muscle', *The Histochemical journal,* 26 (1), 32-40.

Kagey, M. H., et al. (2010), 'Mediator and cohesin connect gene expression and chromatin architecture', *Nature,* 467 (7314), 430-5.

Kanzler, B., et al. (1998), 'Hoxa-2 restricts the chondrogenic domain and inhibits bone formation during development of the branchial area', *Development,* 125 (14), 2587-97.

Katrukha, I. A. (2013), 'Human cardiac troponin complex. Structure and functions', *Biochemistry. Biokhimiia,* 78 (13), 1447-65.

Kim, K. H., et al. (2012), 'Iroquois homeodomain transcription factors in heart development and function', *Circulation research,* 110 (11), 1513-24.

Kind, J. and van Steensel, B. (2010), 'Genome-nuclear lamina interactions and gene regulation', *Current opinion in cell biology,* 22 (3), 320-5.

Klenova, E. M., et al. (2001), 'Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF', *Molecular and cellular biology,* 21 (6), 2221-34.

Krause, A., et al. (2004), 'Tbx5 and Tbx4 transcription factors interact with a new chicken PDZ-LIM protein in limb and heart development', *Developmental biology,* 273 (1), 106-20.

Lee, B. K. and Iyer, V. R. (2012), 'Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation', *The Journal of biological chemistry,* 287 (37), 30906-13.

Lickert, H., et al. (2004), 'Baf60c is essential for function of BAF chromatin remodelling complexes in heart development', *Nature,* 432 (7013), 107-12.

Lobanenkov, V. V., et al. (1990), 'A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene', *Oncogene,* 5 (12), 1743-53.

Lonfat, N. and Duboule, D. (2015), 'Structure, function and evolution of topologically associating domains (TADs) at HOX loci', *FEBS letters,* 589 (20 Pt A), 2869-76.

Luna-Zurita, L. and Bruneau, B. G. (2013), 'Chromatin modulators as facilitating factors in cellular reprogramming', *Current opinion in genetics & development,* 23 (5), 556-61.

Lupianez, D. G., et al. (2015), 'Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions', *Cell,* 161 (5), 1012-25.

Luxan, G., et al. (2016), 'Endocardial Notch Signaling in Cardiac Development and Disease', *Circulation research,* 118 (1), e1-e18.

MacGrogan, D., Nus, M., and de la Pompa, J. L. (2010), 'Notch signaling in cardiac development and disease', *Current topics in developmental biology,* 92, 333-65.

Magdinier, F., Yusufzai, T. M., and Felsenfeld, G. (2004), 'Both CTCF-dependent and -independent insulators are found between the mouse T cell receptor alpha and Dad1 genes', *The Journal of biological chemistry,* 279 (24), 25381-9.

Majumder, P., Gomez, J. A., and Boss, J. M. (2006), 'The human major histocompatibility complex class II HLA-DRB1 and HLA-DQA1 genes are separated by a CTCF-binding enhancer-blocking element', *The Journal of biological chemistry,* 281 (27), 18435-43.

Mark, M., Rijli, F. M., and Chambon, P. (1997), 'Homeobox genes in embryogenesis and pathogenesis', *Pediatric research,* 42 (4), 421-9.

Merkenschlager, M. and Odom, D. T. (2013), 'CTCF and cohesin: linking gene regulatory elements with their targets', *Cell,* 152 (6), 1285-97.

Molto, E., Fernandez, A., and Montoliu, L. (2009), 'Boundaries in vertebrate genomes: different solutions to adequately insulate gene expression domains', *Briefings in functional genomics & proteomics,* 8 (4), 283-96.

Moore, J. M., et al. (2012), 'Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos', *PloS one,* 7 (4), e34915.

Murai, J., et al. (2008), 'Insulation of the ubiquitous Rxrb promoter from the cartilage-specific adjacent gene, Col11a2', *The Journal of biological chemistry,* 283 (41), 27677-87.

Nagy, A., et al. (2003), *Manipulating the mouse embryo: a laboratory manual* (3rd edn.; Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Narendra, V., et al. (2015), 'CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation', *Science,* 347 (6225), 1017-21.

Nasmyth, K. and Haering, C. H. (2009), 'Cohesin: its roles and mechanisms', *Annual review of genetics,* 43, 525-58.

Nikolic, T., et al. (2014), 'The DNA-binding factor Ctcf critically controls gene expression in macrophages', *Cellular & molecular immunology,* 11 (1), 58-70.

Noordermeer, D., et al. (2011), 'The dynamic architecture of Hox gene clusters', *Science,* 334 (6053), 222-5.

Nora, E. P., Dekker, J., and Heard, E. (2013), 'Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods?', *BioEssays : news and reviews in molecular, cellular and developmental biology,* 35 (9), 818-28.

Ong, C. T. and Corces, V. G. (2014), 'CTCF: an architectural protein bridging genome topology and function', *Nature reviews. Genetics,* 15 (4), 234-46.

Peric-Hupkes, D., et al. (2010), 'Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation', *Molecular cell,* 38 (4), 603-13.

Petchey, L. K., et al. (2014), 'Loss of Prox1 in striated muscle causes slow to fast skeletal muscle fiber conversion and dilated cardiomyopathy', *Proceedings of the National Academy of Sciences of the United States of America,* 111 (26), 9515-20.

Peters, T., et al. (2000), 'Organization of mouse Iroquois homeobox genes in two clusters suggests a conserved regulation and function in vertebrate development', *Genome research,* 10 (10), 1453-62.

Phillips-Cremins, J. E., et al. (2013), 'Architectural protein subclasses shape 3D organization of genomes during lineage commitment', *Cell,* 153 (6), 1281-95.

Phillips, J. E. and Corces, V. G. (2009), 'CTCF: master weaver of the genome', *Cell,* 137 (7), 1194-211.

Rada-Iglesias, A., et al. (2011), 'A unique chromatin signature uncovers early developmental enhancers in humans', *Nature,* 470 (7333), 279-83.

Rao, S. S., et al. (2014), 'A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping', *Cell,* 159 (7), 1665-80.

Ribeiro de Almeida, C., et al. (2009), 'Critical role for the transcription regulator CCCTC-binding factor in the control of Th2 cytokine expression', *Journal of immunology,* 182 (2), 999-1010.

Ribeiro de Almeida, C., et al. (2011), 'The DNA-binding protein CTCF limits proximal Vkappa recombination and restricts kappa enhancer interactions to the immunoglobulin kappa light chain locus', *Immunity,* 35 (4), 501-13.

Risca, V. I. and Greenleaf, W. J. (2015), 'Unraveling the 3D genome: genomics tools for multiscale exploration', *Trends in genetics : TIG,* 31 (7), 357-72.

Robinson, M. D. and Oshlack, A. (2010), 'A scaling normalization method for differential expression analysis of RNA-seq data', *Genome biology,* 11 (3), R25.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010), 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics,* 26 (1), 139-40.

Seitan, V. C., et al. (2013), 'Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments', *Genome research,* 23 (12), 2066-77.

Seitan, V. C., et al. (2011), 'A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation', *Nature,* 476 (7361), 467-71.

Seruggia, D. and Montoliu, L. (2014), 'The new CRISPR-Cas system: RNA-guided genome engineering to efficiently produce any desired genetic alteration in animals', *Transgenic research,* 23 (5), 707-16.

Sexton, T. and Cavalli, G. (2015), 'The role of chromosome domains in shaping the functional genome', *Cell,* 160 (6), 1049-59.

Shen, Y., et al. (2012), 'A map of the cis-regulatory sequences in the mouse genome', *Nature,* 488 (7409), 116-20.

Shih, H. Y., et al. (2012), 'Tcra gene recombination is supported by a Tcra enhancer- and CTCF-dependent chromatin hub', *Proceedings of the National Academy of Sciences of the United States of America,* 109 (50), E3493-502.

Smallwood, A. and Ren, B. (2013), 'Genome organization and long-range regulation of gene expression by enhancers', *Current opinion in cell biology,* 25 (3), 387-94.

Smemo, S., et al. (2014), 'Obesity-associated variants within FTO form long-range functional connections with IRX3', *Nature,* 507 (7492), 371-5.

Smith, C. M., et al. (2014), 'The gene expression database for mouse development (GXD): putting developmental expression information at your fingertips', *Developmental dynamics : an official publication of the American Association of Anatomists,* 243 (10), 1176-86.

Solovei, I., Thanisch, K., and Feodorova, Y. (2016), 'How to rule the nucleus: divide et impera', *Current opinion in cell biology,* 40, 47-59.

Soshnikova, N., et al. (2010), 'Functional analysis of CTCF during mammalian limb development', *Developmental cell,* 19 (6), 819-30.

Splinter, E., et al. (2012), 'Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation', *Methods,* 58 (3), 221-30.

Splinter, E., et al. (2006), 'CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus', *Genes & development,* 20 (17), 2349-54.

Stanley, E. G., et al. (2002), 'Efficient Cre-mediated deletion in cardiac progenitor cells conferred by a 3'UTR-ires-Cre allele of the homeobox gene Nkx2-5', *The International journal of developmental biology,* 46 (4), 431-9.

Stennard, F. A. and Harvey, R. P. (2005), 'T-box transcription factors and their roles in regulatory hierarchies in the developing heart', *Development,* 132 (22), 4897-910.

Symmons, O. and Spitz, F. (2013), 'From remote enhancers to gene regulation: charting the genome's regulatory landscapes', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 368 (1620), 20120358.

Takeuchi, J. K., et al. (2011), 'Chromatin remodelling complex dosage modulates transcription factor function in heart development', *Nature communications,* 2, 187.

Tena, J. J., et al. (2011), 'An evolutionarily conserved three-dimensional structure in the vertebrate Irx clusters facilitates enhancer sharing and coregulation', *Nat Commun,* 2, 310.

Trivedi, C. M., et al. (2010), 'Hopx and Hdac2 interact to modulate Gata4 acetylation and embryonic cardiac myocyte proliferation', *Developmental cell,* 19 (3), 450-9.

Valton, A. L. and Dekker, J. (2016), 'TAD disruption as oncogenic driver', *Current opinion in genetics & development,* 36, 34-40.

van de Werken, H. J., et al. (2012a), '4C technology: protocols and data analysis', *Methods in enzymology,* 513, 89-112.

van de Werken, H. J., et al. (2012b), 'Robust 4C-seq data analysis to screen for regulatory DNA interactions', *Nature methods,* 9 (10), 969-72.

van Steensel, B. (2011), 'Chromatin: constructing the big picture', *The EMBO journal,* 30 (10), 1885-95.

van Weerd, J. H., et al. (2014), 'A large permissive regulatory domain exclusively controls Tbx3 expression in the cardiac conduction system', *Circulation research,* 115 (4), 432-41.

Visel, A., et al. (2009), 'ChIP-seq accurately predicts tissue-specific activity of enhancers', *Nature,* 457 (7231), 854-8.

Vostrov, A. A. and Quitschke, W. W. (1997), 'The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation', *The Journal of biological chemistry,* 272 (52), 33353-9.

Vostrov, A. A., Taheny, M. J., and Quitschke, W. W. (2002), 'A region to the N-terminal side of the CTCF zinc finger domain is essential for activating transcription from the amyloid precursor protein promoter', *The Journal of biological chemistry,* 277 (2), 1619-27.

Wallace, J. A. and Felsenfeld, G. (2007), 'We gather together: insulators and genome organization', *Current opinion in genetics & development,* 17 (5), 400-7.

Wan, L. B., et al. (2008), 'Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development', *Development,* 135 (16), 2729-38.

Wang, J., et al. (2015), 'MIR retrotransposon sequences provide insulators to the human genome', *Proceedings of the National Academy of Sciences of the United States of America,* 112 (32), E4428-37.

Wang, Q., et al. (2001), 'Comparative studies on the expression patterns of three troponin T genes during mouse development', *The Anatomical record,* 263 (1), 72-84.

Watson, L. A., et al. (2014), 'Dual effect of CTCF loss on neuroprogenitor differentiation and survival', *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 34 (8), 2860-70.

Wei, E. Q., et al. (2011), 'Fibroblast growth factor homologous factors in the heart: a potential locus for cardiac arrhythmias', *Trends in cardiovascular medicine,* 21 (7), 199-203.

Wilkinson, D. G. and Nieto, M. A. (1993), 'Detection of messenger RNA by *in situ* hybridization to tissue sections and whole mounts', *Methods in enzymology,* 225, 361-73.

Willoughby, D. A., Vilalta, A., and Oshima, R. G. (2000), 'An Alu element from the K18 gene confers position-independent expression in transgenic mice', *The Journal of biological chemistry,* 275 (2), 759-68.

Xie, X., et al. (2007), 'Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites', *Proceedings of the National Academy of Sciences of the United States of America,* 104 (17), 7145-50.

Yang, J. and Corces, V. G. (2012), 'Insulators, long-range interactions, and genome function', *Current opinion in genetics & development,* 22 (2), 86-92.

Yasui, D., et al. (2002), 'SATB1 targets chromatin remodelling to regulate genes over long distances', *Nature,* 419 (6907), 641-5.

Zhang, S. S., et al. (2011), 'Iroquois homeobox gene 3 establishes fast conduction in the cardiac His-Purkinje network', *Proceedings of the National Academy of Sciences of the United States of America,* 108 (33), 13576-81.

Zhong, X. P. and Krangel, M. S. (1999), 'Enhancer-blocking activity within the DNase I hypersensitive site 2 to 6 region between the TCR alpha and Dad1 genes', *Journal of immunology,* 163 (1), 295-300.

Zhu, L., et al. (1995), 'Developmental regulation of troponin I isoform genes in striated muscles of transgenic mice', *Developmental biology,* 169 (2), 487-503.

Zuin, J., et al. (2014), 'Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells', *Proceedings of the National Academy of Sciences of the United States of America,* 111 (3), 996-1001.

APPENDIXES

Appendix A

Paraffin sections RNA *in situ* pre-hybridization buffer

Reagent	Final concentration	Reference
Formamida	50%	Fluka 47670
SSC pH 5,5 20x	5x	
Denhardt´s 100x	2x	
Tween 20	0,1%	Sigma P1379
Chaps	0,1%	Sigma C5070-5G
tRNA	0,05mg/mL	Sigma C5070
Diluted in	H_2O RNAse free	

SSC 20x

Reagent	Final concentration	Reference
NaCl	150mM	Merck 1064041000
Sodium citrate dihydrate	15mM	Sigma C8532
Diluted in	H_2O	

Denhardt´s 100x

Reagent	Final concentration	Reference
Ficoll	0,02 gr/mL	Sigma F2637
polyvinylpyrrolidone	0,02 gr/mL	Sigma P5288
BSA	0,02 gr/mL	Sigma A7906
Dilute in	H_2O RNAse free	

Paraffin sections RNA *in situ* Post-hybridization I buffer

Reagent	Final concentration	Reference
Formamida	50%	
SSC pH 5,5 20x	5x	
SDS	1%	Biorad 161-0418
Diluted in	H_2O RNAse free	

Paraffin sections RNA *in situ* Post-hybridization II buffer

Reagent	Final concentration	Reference
Formamida	50%	
SSC pH 5,5 20x	2x	
SDS	0,2%	Biorad 161-0418
Diluted in	H_2O RNAse free	

Paraffin sections RNA *in situ* Maleic buffer (MABT)

Reagent	Final concentration
MABT 5X	1X
Diluted in	H_2O RNAse free

Paraffin sections RNA *in situ* MABT 5X

Reagent	Final concentration	Reference
Maleic Acid	58,05 gr/L	Sigma M0375
NaCl	43,83 gr/L	Merck 1.06404.1000
Tween 20	0,5%	
Dilute in	H_2O RNAse free	

Autoclave

Paraffin sections RNA *in situ* Blocking solution

Reagent	Final concentration	Reference
Goat serum		Thermofisher C8532
Blocking reagent 10%	1%	
Dilute in	MABT 1X	

Blocking reagent 10% p/v

Reagent	Final concentration	Reference
Blocking reagent	10% (100g in 1L)	Roche 11096176001
Dilute in	MABT 1X	

Paraffin sections RNA *in situ* AP-Buffer

Reagent	Final concentration	Reference
NaCl	0,1 M	
$MgCl_2$	0,05M	Sigma 63072
Tris HCl pH 9,5	0,1 M	Sigma T3253
Tween 20	0,1%	Sigma P9416
Dilute in	H_2O RNAse free	

For embryos from E11.5 add Levamisol (Sigma L9756) at final concentration 2mM

Whole mount RNA *in situ* pre-hybridization buffer

Reagent	Final concentration	Reference
Formamida	50%	
SSC 20x pH 4,5	4x	
tRNA	0,1 gr/mL	
SDS 20%	1%	
Heparin	0.05 gr/mL	
Blocking reagent	1%	
Dilute in	H_2O RNAse free	

Whole mount RNA *in situ* Post-hybridization I buffer

Reagent	Final concentration	Reference
Formamida	50%	
SSC pH 4,5 20x	5x	
SDS	1%	Biorad 161-0418
Diluted in	H_2O RNAse free	

Whole mount RNA *in situ* Post-hybridization II buffer

Reagent	Final concentration	Reference
Formamida	50%	
SSC pH 4,5 20x	2x	
SDS	0,1%	Biorad 161-0418
Diluted in	H_2O RNAse free	

Whole mount RNA *in situ* NTMT Buffer

Reagent	Final concentration	Reference
TrisHCl pH 9,5	0,125 M	
NaCl	0,125 M	
$MgCl_2$	0,0625 M	
Triton X-100 25%	0,125%	Sigma T9284
Levamisol 1M	2,5 mM	
Dilute in	H_2O RNAse free	

TdT buffer 5x

Reagent	Final concentration	Reference
TrisHCl	0.125mM	
Sodium cacodylate trihydrate	0.214gr/mL	Sigma C4945
BSA	0.0012 gr/mL	Sigma A7906
pH	6,6	
Diluted in	H_2O	

TdT buffer 1x

Reagent	Final concentration	Reference
Tdt 5x	1x	
$CoCl_2$	1mM	Sigma 255599
Diluted in	PBS1x	

Prepare at moment of use

Histoblock

Reagent	Final concentration	Reference
BSA	0.03gr/mL	
$MgCl_2$	0.02mM	
Tween 20	3%	Merck 8170721000
Goat serum		
Diluted in	PBS1X	

Microinjection Buffer

Reagent	Final concentration
Tris HCl pH 7,4	50 mM
EDTA	1 mM

Dilute in	Oocyte water

Store at -20ºC

4Cseq Buffer

Reagent	Final concentration	Reference
TrisHCl pH 7,5	50 mM	
NaCl	150 mM	
EDTA	5 mM	
NP-40 (Igepal)	0,5 %	Sigma 13021-100mL
Tritón-100 20%	1%	
Protease inhibitor	1x	Roche 11245200
Diluted in	PBS1x	

Appendix B

Number of lived embryos recovered at different stages.

Age	*Ctcf* $^{fl/+}$	*Ctcf* $^{fl/fl}$	*Ctcf* $^{fl/+}$*; Nkx2.5-Cre*	*Ctcf* $^{fl/fl}$*; Nkx2.5-Cre*
E9.5	26	25	24	24
E10.5	49	48	50	40
E11.5	113	128	109	114
E12.5	5	2	5	6
E13.5	11	2	6	0
E14.5	6	6	4	0
Adult	5	6	7	0

Appendix C

RNAseq data in an excel file in the CD

Appendix D

Table 16. GO terms enriched in all DEG in controls versus *Ctcf* KO comparison

Term	Count	%	P-Value	Fold Enrichment
tetrapyrrole biosynthetic process	12	0.5	7.50E-07	5.7
porphyrin biosynthetic process	12	0.5	7.50E-07	5.7
cellular response to reactive oxygen species	10	0.4	7.00E-05	4.8
cellular response to oxidative stress	11	0.5	7.60E-05	4.3
porphyrin metabolic process	13	0.6	1.20E-05	4.3
tetrapyrrole metabolic process	13	0.6	1.20E-05	4.3
response to reactive oxygen species	15	0.7	2.50E-05	3.6
cell cycle checkpoint	19	0.8	1.20E-05	3.1
translation	108	4.8	8.50E-26	2.9
heart morphogenesis	23	1	2.20E-05	2.7
electron transport chain	33	1.5	8.60E-07	2.5
heart development	64	2.8	8.50E-12	2.5
di-, tri-valent inorganic cation homeostasis	38	1.7	3.10E-06	2.2
cellular di-, tri-valent inorganic cation homeostasis	34	1.5	2.00E-05	2.2
axonogenesis	41	1.8	3.00E-06	2.2
generation of precursor metabolites and energy	65	2.9	3.60E-09	2.1
cell morphogenesis involved in neuron differentiation	44	2	3.70E-06	2.1
neuron projection morphogenesis	42	1.9	9.00E-06	2
cellular cation homeostasis	36	1.6	5.10E-05	2
cation homeostasis	43	1.9	1.20E-05	2
cell projection morphogenesis	47	2.1	5.00E-06	2
cell morphogenesis involved in differentiation	49	2.2	3.70E-06	2
cell part morphogenesis	49	2.2	3.70E-06	2
neuron projection development	50	2.2	3.70E-06	2
blood vessel morphogenesis	45	2	1.50E-05	2
muscle organ development	40	1.8	4.90E-05	2
cellular homeostasis	77	3.4	1.60E-08	1.9
blood vessel development	54	2.4	4.40E-06	1.9
vasculature development	54	2.4	9.30E-06	1.9
cell projection organization	68	3	9.10E-07	1.8
negative regulation of programmed cell death	52	2.3	2.00E-05	1.8
cellular chemical homeostasis	57	2.5	8.20E-06	1.8
negative regulation of cell death	52	2.3	2.30E-05	1.8
ion homeostasis	62	2.8	3.70E-06	1.8
chemical homeostasis	77	3.4	2.50E-07	1.8
cellular ion homeostasis	55	2.4	1.60E-05	1.8
cell morphogenesis	64	2.8	5.40E-06	1.8
tissue morphogenesis	49	2.2	9.00E-05	1.8
tube development	53	2.4	8.90E-05	1.7
homeostatic process	117	5.2	3.00E-09	1.7
cellular component morphogenesis	70	3.1	7.30E-06	1.7
cytoskeleton organization	65	2.9	1.60E-05	1.7
phosphorus metabolic process	157	7	7.90E-09	1.6
phosphate metabolic process	157	7	7.90E-09	1.6
phosphorylation	129	5.7	3.40E-07	1.5

Table 17. GO terms enriched in downregulated DEG in controls versus *Ctcf* KO comparison

Term	Count	%	P-Value	Fold Enrichment
heart morphogenesis	22	1.9	1.30E-09	4.9
cardiac muscle tissue development	15	1.3	4.80E-06	4.4
regulation of heart contraction	13	1.1	3.00E-05	4.3
heart development	57	4.8	9.10E-21	4.3
actin filament organization	13	1.1	9.80E-05	3.9
transmembrane receptor protein serine/threonine kinase signaling pathway	17	1.4	1.30E-05	3.6
axon guidance	21	1.8	1.20E-06	3.6
cartilage development	16	1.4	5.20E-05	3.4
axonogenesis	33	2.8	2.10E-09	3.4
regulation of MAP kinase activity	17	1.4	3.40E-05	3.4
actin cytoskeleton organization	33	2.8	2.90E-09	3.3
actin filament-based process	35	3	1.00E-09	3.3
regulation of cell migration	18	1.5	2.90E-05	3.3
branching morphogenesis of a tube	18	1.5	3.40E-05	3.2
neuron projection morphogenesis	34	2.9	4.10E-09	3.2
tube morphogenesis	33	2.8	7.40E-09	3.2
cell morphogenesis involved in neuron differentiation	35	3	2.60E-09	3.2
striated muscle tissue development	24	2	1.70E-06	3.1
cell morphogenesis involved in differentiation	40	3.4	2.80E-10	3.1
cell projection morphogenesis	38	3.2	9.20E-10	3.1
blood vessel morphogenesis	37	3.1	1.90E-09	3.1
blood vessel development	45	3.8	4.00E-11	3.1
morphogenesis of a branching structure	23	1.9	4.70E-06	3.1
cell part morphogenesis	39	3.3	1.00E-09	3.1
neuron projection development	40	3.4	6.60E-10	3.1
muscle organ development	32	2.7	5.50E-08	3
vasculature development	45	3.8	9.10E-11	3
regulation of cell motion	19	1.6	6.30E-05	3
muscle tissue development	24	2	5.80E-06	2.9
muscle cell differentiation	20	1.7	6.50E-05	2.8
cell projection organization	54	4.6	1.00E-11	2.8
tissue morphogenesis	40	3.4	8.80E-09	2.8
tube development	44	3.7	1.80E-09	2.8
enzyme linked receptor protein signaling pathway	45	3.8	1.70E-09	2.7
cell morphogenesis	50	4.2	3.50E-10	2.7
extracellular structure organization	23	1.9	7.90E-05	2.6
cellular component morphogenesis	54	4.6	4.10E-10	2.6
transmembrane receptor protein tyrosine kinase signaling pathway	29	2.4	1.10E-05	2.5
regulation of kinase activity	29	2.4	1.10E-05	2.5
regulation of protein kinase activity	28	2.4	1.70E-05	2.5
morphogenesis of an epithelium	26	2.2	3.80E-05	2.5
neuron development	43	3.6	1.10E-07	2.4
regulation of transferase activity	29	2.4	2.20E-05	2.4
regulation of system process	28	2.4	7.00E-05	2.3
cytoskeleton organization	45	3.8	3.60E-07	2.3
regulation of phosphorylation	39	3.3	4.60E-06	2.2
chemical homeostasis	48	4.1	5.70E-07	2.2
chromatin modification	31	2.6	8.00E-05	2.2
cell motion	48	4.1	6.70E-07	2.2
regulation of phosphorus metabolic process	39	3.3	1.10E-05	2.2
regulation of phosphate metabolic process	39	3.3	1.10E-05	2.2
embryonic morphogenesis	46	3.9	2.10E-06	2.1
protein amino acid phosphorylation	82	6.9	9.20E-11	2.1
in utero embryonic development	34	2.9	6.30E-05	2.1
ion homeostasis	37	3.1	3.40E-05	2.1
phosphate metabolic process	109	9.2	1.20E-13	2.1
phosphorus metabolic process	109	9.2	1.20E-13	2.1
phosphorylation	89	7.5	7.30E-11	2.1
neuron differentiation	48	4.1	7.00E-06	2
embryonic development ending in birth or egg hatching	50	4.2	8.10E-06	2
cellular homeostasis	40	3.4	9.10E-05	1.9
chordate embryonic development	48	4.1	2.80E-05	1.9
cell adhesion	63	5.3	2.10E-06	1.9
biological adhesion	63	5.3	2.30E-06	1.9
homeostatic process	63	5.3	7.80E-06	1.8
intracellular signaling cascade	93	7.9	5.20E-07	1.7

Table 18. GO terms enriched in upegulated DEG in controls versus *Ctcf* KO comparison

Term	Count	%	P-Value	Fold Enrichment
tetrapyrrole biosynthetic process	10	0.9	1.70E-07	9.8
porphyrin biosynthetic process	10	0.9	1.70E-07	9.8
tetrapyrrole metabolic process	11	1	7.10E-07	7.5
porphyrin metabolic process	11	1	7.10E-07	7.5
response to reactive oxygen species	12	1.1	2.80E-06	5.9
translation	104	9.7	2.90E-51	5.8
oxidative phosphorylation	15	1.4	1.80E-06	4.7
electron transport chain	29	2.7	1.40E-11	4.6
erythrocyte homeostasis	13	1.2	2.00E-05	4.5
heterocycle biosynthetic process	13	1.2	2.00E-05	4.5
erythrocyte differentiation	12	1.1	5.60E-05	4.4
cofactor biosynthetic process	18	1.7	1.10E-05	3.5
generation of precursor metabolites and energy	45	4.2	5.00E-11	3.1
oxidation reduction	70	6.5	8.10E-07	1.8

Table 19. GO terms enriched in all DEG in heterozygote versus *Ctcf* KO comparison

Term	Count	%	P-Value	Benjamini
translation	103	5.3	1.10E-27	2.10E-24
cellular protein metabolic process	318	16.5	8.00E-16	7.40E-13
heart development	55	2.9	7.10E-10	4.60E-07
ion homeostasis	62	3.2	2.90E-08	1.40E-05
cellular chemical homeostasis	56	2.9	2.30E-07	8.90E-05
striated muscle tissue development	33	1.7	8.20E-07	2.60E-04
cardiac muscle tissue development	20	1	1.60E-06	4.50E-04
tetrapyrrole biosynthetic process	11	0.6	2.20E-06	5.40E-04
porphyrin biosynthetic process	11	0.6	2.20E-06	5.40E-04
neuron projection development	46	2.4	2.40E-06	5.20E-04
muscle tissue development	33	1.7	4.00E-06	7.70E-04
regulation of heart contraction	18	0.9	4.30E-06	7.40E-04
muscle organ development	39	2	4.80E-06	7.70E-04
striated muscle cell differentiation	25	1.3	5.40E-06	7.90E-04
heart morphogenesis	22	1.1	8.70E-06	1.20E-03
cell morphogenesis involved in neuron differentiation	39	2	1.10E-05	1.40E-03
hexose metabolic process	37	1.9	1.20E-05	1.40E-03
axonogenesis	36	1.9	1.30E-05	1.40E-03
purine nucleoside triphosphate metabolic process	27	1.4	1.40E-05	1.50E-03
cell part morphogenesis	43	2.2	1.50E-05	1.50E-03
cell morphogenesis involved in differentiation	43	2.2	1.50E-05	1.50E-03
cellular cation homeostasis	34	1.8	1.80E-05	1.70E-03
porphyrin metabolic process	12	0.6	2.00E-05	1.80E-03
erythrocyte homeostasis	17	0.9	2.60E-05	2.20E-03
tissue morphogenesis	46	2.4	2.60E-05	2.20E-03
neuron projection morphogenesis	37	1.9	3.00E-05	2.40E-03
tissue development	97	5	4.00E-05	3.00E-03
erythrocyte differentiation	16	0.8	4.80E-05	3.60E-03
blood vessel development	46	2.4	5.00E-05	3.50E-03
glucose metabolic process	31	1.6	5.40E-05	3.70E-03
cell projection morphogenesis	40	2.1	5.50E-05	3.60E-03
phosphorylation	106	5.5	5.60E-05	3.50E-03
nucleoside phosphate metabolic process	46	2.4	6.70E-05	4.10E-03
muscle cell development	18	0.9	7.90E-05	4.70E-03
cell morphogenesis	54	2.8	8.50E-05	4.90E-03
vasculature development	46	2.4	9.00E-05	5.10E-03
regulation of protein metabolic process	59	3.1	1.20E-04	6.30E-03
purine ribonucleotide metabolic process	27	1.4	1.20E-04	6.30E-03
blood vessel morphogenesis	38	2	1.70E-04	8.80E-03
regulation of striated muscle tissue development	14	0.7	1.70E-04	8.60E-03
cellular polysaccharide metabolic process	14	0.7	2.20E-04	1.10E-02
interphase of mitotic cell cycle	15	0.8	2.50E-04	1.20E-02
glucan metabolic process	12	0.6	2.60E-04	1.20E-02
cellular glucan metabolic process	12	0.6	2.60E-04	1.20E-02
interphase	15	0.8	3.90E-04	1.80E-02
in utero embryonic development	46	2.4	4.20E-04	1.90E-02
nucleobase, nucleoside and nucleotide biosynthetic process	34	1.8	4.80E-04	2.10E-02
nucleobase, nucleoside, nucleotide and nucleic acid biosynt	34	1.8	4.80E-04	2.10E-02
neuron development	49	2.5	4.90E-04	2.10E-02
purine nucleotide biosynthetic process	28	1.5	5.30E-04	2.20E-02
striated muscle cell development	15	0.8	7.40E-04	3.00E-02
energy reserve metabolic process	12	0.6	7.90E-04	3.10E-02
negative regulation of programmed cell death	42	2.2	7.90E-04	3.00E-02
negative regulation of cell death	42	2.2	8.60E-04	3.20E-02
regulation of phosphate metabolic process	49	2.5	9.80E-04	3.60E-02
regulation of phosphorus metabolic process	49	2.5	9.80E-04	3.60E-02

Table 20. GO terms enriched in downregulated DEG in heterozygote versus *Ctcf* KO comparison

Term	Count	%	P-Value	Benjamini
heart development	49	5.1	5.40E-19	8.10E-16
tissue development	74	7.7	2.70E-12	2.10E-09
regulation of heart contraction	18	1.9	7.20E-11	3.60E-08
cardiac muscle tissue development	19	2	8.00E-11	3.00E-08
heart morphogenesis	21	2.2	1.70E-10	5.10E-08
striated muscle tissue development	27	2.8	2.70E-10	6.70E-08
muscle organ development	32	3.3	2.80E-10	6.00E-08
tissue morphogenesis	37	3.9	8.40E-10	1.60E-07
muscle tissue development	27	2.8	1.30E-09	2.10E-07
striated muscle cell differentiation	21	2.2	5.70E-09	8.70E-07
blood vessel development	36	3.8	6.10E-09	8.30E-07
vasculature development	36	3.8	1.20E-08	1.50E-06
neuron projection development	33	3.4	1.60E-08	1.80E-06
cell morphogenesis involved in differentiation	32	3.3	2.90E-08	3.10E-06
organ morphogenesis	61	6.4	3.30E-08	3.40E-06
enzyme linked receptor protein signaling pathway	37	3.9	3.50E-08	3.30E-06
axonogenesis	27	2.8	6.50E-08	5.70E-06
cell part morphogenesis	31	3.2	1.00E-07	8.60E-06
cell morphogenesis involved in neuron differentiation	28	2.9	1.70E-07	1.30E-05
ion homeostasis	37	3.9	2.10E-07	1.60E-05
blood vessel morphogenesis	29	3	2.80E-07	2.00E-05
cell morphogenesis	38	4	2.80E-07	1.90E-05
neuron projection morphogenesis	27	2.8	3.10E-07	2.00E-05
cell projection morphogenesis	29	3	4.20E-07	2.60E-05
neurogenesis	54	5.6	4.60E-07	2.80E-05
generation of neurons	51	5.3	5.70E-07	3.30E-05
muscle cell development	15	1.6	8.50E-07	4.80E-05
regulation of cell migration	18	1.9	1.50E-06	8.30E-05
cellular chemical homeostasis	33	3.4	1.90E-06	9.90E-05
cardiac cell differentiation	11	1.1	2.10E-06	1.10E-04
neuron development	34	3.6	4.40E-06	2.20E-04
phosphorylation	63	6.6	4.70E-06	2.20E-04
regulation of signal transduction	59	6.2	5.90E-06	2.70E-04
glucan metabolic process	10	1	1.80E-05	7.90E-04
cellular glucan metabolic process	10	1	1.80E-05	7.90E-04
hexose metabolic process	23	2.4	2.00E-05	8.60E-04
regulation of phosphorus metabolic process	33	3.4	2.20E-05	9.10E-04
regulation of phosphate metabolic process	33	3.4	2.20E-05	9.10E-04
regulation of striated muscle tissue development	11	1.1	2.30E-05	9.20E-04
gland development	25	2.6	2.60E-05	1.00E-03
cellular polysaccharide metabolic process	11	1.1	2.80E-05	1.10E-03
in utero embryonic development	30	3.1	3.50E-05	1.30E-03
glucose metabolic process	20	2.1	4.10E-05	1.50E-03
striated muscle cell development	12	1.3	4.30E-05	1.50E-03
energy reserve metabolic process	10	1	4.80E-05	1.70E-03
chordate embryonic development	40	4.2	6.70E-05	2.30E-03
regulation of metal ion transport	11	1.1	7.70E-05	2.60E-03
morphogenesis of an epithelium	22	2.3	8.70E-05	2.90E-03
protein modification process	90	9.4	1.20E-04	3.80E-03
branching morphogenesis of a tube	15	1.6	1.40E-04	4.30E-03
regulation of kinase activity	23	2.4	1.40E-04	4.30E-03
regulation of ion transport	12	1.3	1.60E-04	4.80E-03
gland morphogenesis	14	1.5	1.80E-04	5.20E-03
epithelium development	28	2.9	2.80E-04	8.00E-03
cellular cation homeostasis	19	2	3.70E-04	1.10E-02
positive regulation of cell proliferation	28	2.9	5.80E-04	1.60E-02
regulation of cell development	19	2	6.40E-04	1.70E-02
intracellular signaling cascade	66	6.9	9.30E-04	2.50E-02

Table 21. GO terms enriched in upegulated DEG in controls versus *Ctcf* KO comparison, part I

Term	Count	%	P-Value	Benjamini
translation	97	10	3.30E-47	3.80E-44
cellular protein metabolic process	201	20.7	4.80E-19	2.80E-16
tetrapyrrole biosynthetic process	11	1.1	5.00E-09	2.00E-06
porphyrin biosynthetic process	11	1.1	5.00E-09	2.00E-06
porphyrin metabolic process	12	1.2	3.30E-08	9.70E-06
erythrocyte homeostasis	16	1.6	3.50E-08	8.30E-06
erythrocyte differentiation	15	1.5	1.10E-07	2.20E-05
heme biosynthetic process	8	0.8	1.50E-06	2.50E-04
heme metabolic process	9	0.9	2.80E-06	4.00E-04
nucleoside phosphate metabolic process	31	3.2	2.00E-05	2.60E-03
cellular response to reactive oxygen species	8	0.8	2.20E-05	2.60E-03
nucleobase, nucleoside and nucleotide biosynthetic process	24	2.5	7.70E-05	8.10E-03
nucleobase, nucleoside, nucleotide and nucleic acid biosynthetic p	24	2.5	7.70E-05	8.10E-03
erythrocyte development	7	0.7	1.10E-04	1.10E-02
purine nucleoside triphosphate metabolic process	17	1.7	1.40E-04	1.30E-02
oxygen transport	6	0.6	1.50E-04	1.20E-02
myeloid cell differentiation	16	1.6	1.60E-04	1.20E-02
hydrogen peroxide catabolic process	6	0.6	2.40E-04	1.80E-02
pyrimidine nucleotide biosynthetic process	7	0.7	4.50E-04	3.10E-02
nucleoside triphosphate biosynthetic process	15	1.5	4.70E-04	3.00E-02
purine ribonucleotide metabolic process	17	1.7	5.50E-04	3.40E-02
ribonucleotide biosynthetic process	16	1.6	7.70E-04	4.40E-02
purine nucleotide biosynthetic process	18	1.8	9.50E-04	5.20E-02
response to hydrogen peroxide	7	0.7	1.30E-03	6.70E-02
proton transport	10	1	1.50E-03	7.50E-02
regulation of protein metabolic process	33	3.4	2.00E-03	9.40E-02
negative regulation of programmed cell death	25	2.6	2.70E-03	1.20E-01
negative regulation of cell death	25	2.6	2.90E-03	1.20E-01
ATP synthesis coupled proton transport	8	0.8	3.00E-03	1.20E-01
energy coupled proton transport, down electrochemical gradient	8	0.8	3.00E-03	1.20E-01
translational initiation	8	0.8	3.50E-03	1.40E-01
hemopoiesis	25	2.6	3.90E-03	1.50E-01
negative regulation of apoptosis	24	2.5	4.30E-03	1.60E-01
protein targeting to mitochondrion	6	0.6	5.20E-03	1.80E-01
protein localization in organelle	13	1.3	5.20E-03	1.80E-01
ion transmembrane transport	8	0.8	6.20E-03	2.00E-01
mitochondrial transport	8	0.8	6.20E-03	2.00E-01
hemopoietic or lymphoid organ development	26	2.7	8.20E-03	2.50E-01
protein transport	50	5.1	8.70E-03	2.50E-01
negative regulation of actin filament polymerization	6	0.6	9.20E-03	2.60E-01
protein targeting	15	1.5	1.10E-02	3.00E-01
protein import	11	1.1	1.10E-02	3.00E-01
intracellular protein transport	25	2.6	1.20E-02	3.10E-01
negative regulation of protein polymerization	6	0.6	1.30E-02	3.10E-01
negative regulation of protein complex assembly	6	0.6	1.30E-02	3.10E-01
interphase of mitotic cell cycle	8	0.8	1.40E-02	3.40E-01
regulation of actin polymerization or depolymerization	8	0.8	1.60E-02	3.60E-01
regulation of translation	12	1.2	1.70E-02	3.70E-01
regulation of actin filament length	8	0.8	1.80E-02	3.80E-01
interphase	8	0.8	1.80E-02	3.80E-01
anti-apoptosis	11	1.1	1.80E-02	3.80E-01
protein catabolic process	42	4.3	2.10E-02	4.20E-01
ion homeostasis	25	2.6	2.40E-02	4.50E-01
serine family amino acid metabolic process	5	0.5	2.70E-02	4.90E-01
cellular amino acid biosynthetic process	7	0.7	2.80E-02	4.90E-01
cellular chemical homeostasis	23	2.4	2.90E-02	4.90E-01
regulation of protein polymerization	8	0.8	3.10E-02	5.10E-01
regulation of actin filament polymerization	7	0.7	3.10E-02	5.10E-01

Table 22. GO terms enriched in upegulated DEG in controls versus *Ctcf* KO comparison, part II

cellular cation homeostasis	15	1.5	3.20E-02	5.10E-01
regulation of actin filament depolymerization	5	0.5	3.60E-02	5.50E-01
regulation of actin cytoskeleton organization	8	0.8	3.90E-02	5.70E-01
ncRNA metabolic process	18	1.8	4.10E-02	5.90E-01
ncRNA processing	15	1.5	4.20E-02	5.90E-01
regulation of cellular protein metabolic process	23	2.4	4.40E-02	5.90E-01
cellular protein catabolic process	39	4	4.40E-02	5.90E-01
cell redox homeostasis	8	0.8	4.60E-02	6.00E-01
negative regulation of neuron apoptosis	7	0.7	4.80E-02	6.10E-01
protein import into mitochondrial inner membrane	3	0.3	5.00E-02	6.20E-01
erythrocyte maturation	3	0.3	5.00E-02	6.20E-01
coenzyme biosynthetic process	8	0.8	5.30E-02	6.30E-01
M phase of mitotic cell cycle	17	1.7	5.40E-02	6.40E-01
G1/S transition of mitotic cell cycle	5	0.5	5.90E-02	6.70E-01
protein homooligomerization	7	0.7	6.10E-02	6.70E-01
RNA processing	32	3.3	6.20E-02	6.80E-01
inner mitochondrial membrane organization	3	0.3	6.40E-02	6.80E-01
ATP synthesis coupled electron transport	4	0.4	6.70E-02	6.90E-01
RNA metabolic process	45	4.6	7.50E-02	7.30E-01
translational elongation	6	0.6	7.70E-02	7.30E-01
actin filament capping	4	0.4	7.70E-02	7.30E-01
neurofilament cytoskeleton organization	3	0.3	7.90E-02	7.30E-01
amine biosynthetic process	8	0.8	9.30E-02	7.90E-01
metallo-sulfur cluster assembly	3	0.3	9.60E-02	7.90E-01
maternal placenta development	3	0.3	9.60E-02	7.90E-01
hemoglobin metabolic process	3	0.3	9.60E-02	7.90E-01
posttranscriptional regulation of gene expression	13	1.3	9.60E-02	7.90E-01
negative regulation of actin filament depolymerization	4	0.4	9.80E-02	7.90E-01
modification-dependent macromolecule catabolic process	35	3.6	9.80E-02	7.90E-01
regulation of programmed cell death	38	3.9	1.00E-01	7.90E-01

Appendix E

Manuscript of this work in its current version and published article as a result of a collaboration are included in the CD.